TOPICS IN NEUTRINO PHYSICS AND COSMOLOGY

by

Louis Anthony Lello

B.S. in Chemistry, Mathematics, Eastern Michigan University, 2010

M.S. in Physics, University of Pittsburgh, 2012

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Louis Anthony Lello

It was defended on

September 27, 2016

and approved by

Daniel Boyanovsky, Professor, Department of Physics and Astronomy

Richard Holman, Professor, Department of Physics

Adam Leibovich, Professor, Department of Physics and Astronomy

Vittorio Paolone, Professor, Department of Physics and Astronomy

Xiao-Lun Wu, Professor, Department of Physics and Astronomy

Dissertation Director: Daniel Boyanovsky, Professor, Department of Physics and Astronomy

TOPICS IN NEUTRINO PHYSICS AND COSMOLOGY

Louis Anthony Lello, PhD

University of Pittsburgh, 2016

This thesis focuses on timely issues in particle physics of neutrino phenomenology and cosmology. In this work, the methods of quantum optics, quantum field theory in curved spacetime and non-equilibrium field theory are employed to study the issues of inflation and neutrino mixing and oscillations in terrestrial and cosmological settings. Anomalies in short baseline oscillation experiments have provided hints of an additional neutrino species which are used as motivation to study potential experimental signatures through decays at rest, modifications to the treatment of appearance/disappearance oscillation experiments and experiments searching for lepton number violation. The role of decoherence and entanglement in the production mechanism of a two body decay is explored further in Minkowski and de Sitter spacetime with the aim of studying correlations in the CMB. This work is extended in Minkowski spacetime by studying decoherence issues in the cascade decay of successive two body decays with the aim of searching for new particles. The effect of initial conditions for inflation is explored with the motivation of persistent large scale anomalies and with an eye towards the soon-to-be-measured tensor to scalar ratio. Inclusion of these non-trivial initial conditions introduces large scale power suppression and potentially observable oscillations into the tensor to scalar ratio. The issue of dark matter is explored through the possibility of previously unstudied sources a sterile neutrino dark matter candidate. Pions, which are the most abundant particle after the QCD phase transition, can lead to the production of sterile neutrinos which yields a non negligible abundance of dark matter compared to established mechanisms in the literature. This additional source of sterile neutrino naturally leads to a multi component distribution function and a natural scenario for mixed dark matter. Further

sources of sterile neutrinos are discussed and the possibility of sterile neutrinos produced at electroweak temperatures is explored and leads to a generalized temperature dependence of active sterile mixing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Firstly, I would like to thank my PhD advisor, Dr. Daniel Boyanovsky. It has been a unique privilege to have been able to learn from and work with Daniel, he has been the single most influential and effective teacher from which I've had the pleasure of learning. His youthful enthusiasm for physics and seemingly limitless patience as a teacher have made my time in Pittsburgh precious and productive. This enthusiasm for learning spreads beyond the realm of physics and extends to any subject that beauty may be found - an enthusiasm which he fosters in those around him. I consider myself quite lucky to have found a mentor whose love of learning remains undiminished in spite of the highs and lows associated with many years academia. My gratitude towards Dr. Boyanovsky cannot be overstated and I offer him my most heartfelt thanks.

I would like to extend my thanks to the members of my PhD committee: Rich Holman, Adam Leibovich, Vittorio Paolone, Xiao-Lun Wu and Andrew Daley. They have provided many thoughtful questions and helpful suggestions throughout the course of my time as a graduate student. I would particularly like to thank Xiao-Lun Wu for stepping in at the last stage to cover a difficult scheduling conflict.

I would like to thank Eric Swanson, Rich Holman and Ira Rothstein for teaching me a great deal of quantum field theory and particle physics. Their courses provided a stimulating and enjoyable learning environment.

I thank Rob Pisarski and the Nuclear Theory group at Brookhaven National Laboratory for their hospitality during my stay.

I would also like to thank my collaborators, Richard Holman and Robert Pisarski for additional training and collaboration.

A special thank you is owed to Leyla Hirschfeld for the uncountably many bureaucratic hurdles she has helped to overcome. She has unequivocally made the process of navigating grad school much more manageable and enjoyable.

The difficulty of coping with life in graduate school was made far more tolerable from the support of friends and family. I would like to thank my family for their support and encouragement. My friends and fellow graduate students deserve thanks for their camaraderie and commiseration; in particular, I would like to thank Tim Licquia, Alex Maries, Kevin Sapp, Travis Hurst and Daniel Wiegand for the countless physics discussions, escapades and laughs.

This work was supported in part by the Andrew Mellon Predoctoral Fellowship, the University of Pittsburgh Arts and Sciences Predoctoral Fellowship and the PITT-PACC Predoctoral Fellowship. This work received additional support through the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under contract number DEAC05-06OR23100. Partial support for this thesis was also received from the NSF through the grants PHY-0852497, PHY-1202227, PHY-1506912 and through the U.S. Department of Energy under contract DE-SC0012704.

1.0 INTRODUCTION

This first chapter is to provide an overview of the current status of the Standard Models of particle physics and cosmology while describing the clearest immediate issues of observational and experimental interest which remain conclusively unresolved. Explanations of how these issues may be resolved will be presented and serve as a motivation for the work of this thesis. After setting the background and motivations, a brief summary of the work is presented.

Chapters 2-11 each provide a basic introduction and delve deeper into the appropriate background for the topic before the technical details are discussed. Readers who are uninterested in the technical details are encouraged to read this introduction section and the summary of results. Readers with a higher level of interest are encouraged to read the body of a particular chapter for technical details and more in depth discussion. Chapter 12 provides a summary of the main results of the previous chapters.

A note to the interested reader: There is a level of redundancy in several of the chapters. For instance, the Wigner-Weisskopf method is discussed at length in chapters 2,4,5,7 however the application and implementation of these sections is unique to a particular chapter. This redundancy is tolerated and required so that each chapter may stand alone as an independent piece of work for the reader who may only be interested in particular chapters.

1.1 THE STANDARD MODELS OF PARTICLE PHYSICS AND COSMOLOGY

The Standard Model of Particle Physics was created and refined in 1960-1980 by combining many deep insights of the past years - the principle of gauge symmetry, spontaneous symmetry breaking, electroweak unification [4, 5, 6, 7]. This model explained the observations from years of collider physics and successfully predicted additional particles which drove experimental efforts for years. Recently, the final missing particle of the standard model was found by the CMS/Atlas detectors at CERN's Large Hadron Collider [2, 3]- a particle we refer to as the Higg's particle [7]. This final particle was sought for many years as its existence is the conceptual bedrock of the standard model [8]. Through spontaneous symmetry breaking, the Higgs potential provides a mechanism to provide a mass to heavy vector bosons - mass terms which would be otherwise unallowed by gauge invariance [9]. Its discovery was a great success of the pioneering work of the late 20th century physics.

The Standard Model is built off of two main sectors, the electroweak interactions and the strong interactions. The strong interaction is the force which holds nucleons together and was formulated by identifying patterns through the use of group theory on the particle zoo of mesons and baryons [10]. The electroweak sector required several deep insights and for which the Higgs particle is needed. The electroweak sector had its beginnings in the process of beta decay, where a neutron decays to a proton, electron and antineutrino [11]. Fermi introduced a point interaction to describe this process but it was quickly realized that this theory broke down at large energies [12, 13]. It was discovered that there must be another photon like particle responsible for this process but must have a large, nonzero mass. It was also shown in a brilliant experiment that this weak interaction was special in that it violated the symmetries of charge conjugation and parity - symmetries that were previously taken for granted [14, 15, 16].

It was understood that the interaction must be a local interaction and this principle guided the implementation of the observed symmetry for this new force. Theories that obey a locality principle under a symmetry are known as gauge theories, the symmetry itself being

called a gauge symmetry. This gauge symmetry was present in the basic theories of quantum electrodynamics where the electrons serve as the matter particles and the massless photon is the gauge particle which is responsible for the interaction [17, 18]. The gauge symmetry principle showed itself to be quite instrumental in constructing theories that agreed with experiment and soon was realized to be a fundamental guiding principle. The new massive gauge particle from the electroweak theory proved problematic in the sense that its existence could not be reconciled with the principle of gauge invariance - a massive photon like particle explicitly broke the gauge symmetry [9].

The insight was to borrow an idea found in the phenomena of superconductivity where electrons bind into pairs and develop a condensate - this condensate leads to the massless photon of the theory developing a nonzero mass. In the electroweak theory, a new particle was introduced, soon to be known as the Higgs particle, which featured a potential that would lead to nonzero vacuum expectation value. This particle is then coupled to massless gauge particles of the electroweak sector through the gauge symmetry and, as this new field is expanded around its non zero vacuum expectation value, several gauge bosons acquired a nonzero mass while one remained massless - these were the massive W^+, W^-, Z^0 and the massless photon [8, 17, 18, 9]. The mechanism of spontaneous symmetry breaking also gave masses to the fermionic matter particles of the standard model through Yukawa interactions with the Higgs. This has been known since circa 1970 and the discovery of the Higgs didn't happen until 2012 [2, 3]. When the strong and electroweak forces were combined together with the spontaneously broken Higgs field, the Standard Model of Particle Physics was born.

The Standard Model of Particle Physics provides a highly accurate theoretical framework which is consistent with a wide array of experimental observations of phenomena down to the femtometer distance scale and the limits of applicability of this model are currently unknown. On the other extreme distance scale, the concordance ΛCDM model is the standard model for cosmology - describing phenomena over distances of billions of light years. This model is deeply rooted in the principles of general relativity and is based on the several crucial observations - the expansion of the universe, the cosmic microwave background radiation and big bang nucleosynthesis. The core of this model is known as big bang cosmology and this core is modified to explain several discrepancies with observations [317].

The expansion of the universe was discovered by Edwin Hubble and the expansion is described by the appropriately names Hubble's law. Hubble used cepheids, a pulsating star whose precise intrinsic luminosity-oscillation period relation allows the use as standard candles, to show that there exists a relation between the speed of recession and the distance for astrophysical bodies [1]. This observation established that universe was expanding - counter to the belief of a static, eternal universe. The observation by Hubble of the cosmic expansion, combined with the machinery of general relativity, yields a profound understanding of large scale evolution of the cosmos.

The simple observational fact that the universe is expanding leads to important predictions - most notably that of the cosmic microwave background radiation. It is fairly easy to understand how this background radiation occurs with the right picture in mind. As the universe expands, any two given points will be stretched further apart as time moves forward and, conversely, the points will be closer at times in the past. A visual analogy can be made with insects standing still on the surface of a balloon - as the balloon expands, an insect will move further from any other insect but, if the balloon is deflated, the insects will be closer to one another. In this example, the density of insects decreases as the balloon expands while the density of insects increases with smaller balloon size.

In cosmology, the same basic situation occurs where neutral hydrogen atoms play the role of the insects while spacetime itself plays the role of the balloon. In the past, the neutral hydrogen atoms are packed together very tightly and form a high temperature plasma - temperatures so high that the hydrogen atoms are broken into its constituents, free nucleons and electrons. In this nucleon/electron plasma, photons undergo rapid collisions with the charged particles of the plasma - collisions that are explicitly due to the presence of free protons, electrons and other charged particles. With this photon trapping fluid formed in the past, the universe expands and cools as time moves forward. After cooling enough, the plasma will not have enough energy to keep the nucleons and electrons from binding into neutral atoms. When this neutral hydrogen fluid is formed, as there are no charged particles, the photons no longer scatter rapidly and propagate through space unimpeded.

From the expansion of the universe, we are lead to the picture that photons were bound in

a nucleon/electron fluid which then cooled and coalesced into a neutral hydrogen fluid. The hydrogen fluid is less effective at scattering photons and increases the mean free path of the photons - this is the photons "last scattering" with the plasma. The photons are decoupled from the hydrogen plasma and travel freely through the universe until they strike an opaque object, like a detector on earth. The expansion of the universe predicts the existence of this last scattering of light distributed homogenously across the sky. Serendipitously in 1964, Penzias and Wilson accidentally discovered this radiation while studying radio waves from balloon experiments. This is the cosmic microwave background raditaion (CMB) which has been mapped extensively by the COBE, WMAP and Planck satellites [247, 63, 317].

With this picture in mind and confirmation through the CMB, the next obvious step is to study the nuclear chemical reactions in this primeval hot soup. If this picture is accurate then all observed matter should be a descendant of this hot big bang state and, given a knowledge of nuclear reactions, a prediction of the relative amount of protons, neutrons, photons and other light elements should be possible. The main parameters in the theory of the chain of nuclear reactions are the proton to neutron and photon to baryon ratios. Using known nuclear chemical reactions and statistical mechanics, it is possible to make a prediction of the relative abundance of light elements [438, 439, 440, 437, 441]. There are various techniques to measure primeval nuclear abundances of various elements, such as ${}^2H, {}^3He, {}^4He, {}^7Li$, where the match between observations and predictions from theory are striking.

The expansion of the universe, the CMB and BBN form the bedrock of the standard big bang cosmology. This standard frame work only needs augmentation with the observations of dark matter, dark energy and the paradigm of inflation to form the concordance ΛCDM model, a model which explains most of the observations in cosmology today. The issues of dark matter and inflation arise from problems in the big bang cosmology and reflect an incomplete picture of the Standard Model of particle physics. The theoretical and observational work of the past century has ushered in an exciting era where the problems on high energy, short-distance scales have merged with the problems on vast cosmological, long-distance scales.

1.2 SHORTCOMINGS OF THE MODELS AND MOTIVATION FOR EXTENSION

Despite the astounding successes of the past century, there remains several observational and experimental results which are not accommodated for in these standard models. Cosmological observations of distant galaxies and the CMB have indicated that there exists a ubiquitous and vast excess of matter in astrophysical environments which is completely unaccounted for in the Standard Model of particle physics - this is known as dark matter [317, 318, 319, 320]. In particle physics, the neutrinos, a weakly interacting particle, are described in the Standard Model as massless objects; however, it has been found experimentally that neutrinos possess a non-zero mass and oscillate between the different flavors of neutrinos [26, 27, 28, 29]. This phenomena has spurred a very active and exciting experimental effort, yet these properties of neutrinos remain absent in the Standard Model. On the cosmological front, the precision observations of the CMB have lead to the confirmation that the universe is homogenous and isotropic to one part in 100000 [247, 63]. This level of homogeneity is beyond what one would expect from the size of causally disconnected regions and requires a new mechanism - the commonly accepted explanation is an additional stage of cosmic evolution called inflation [237, 238, 239, 240]. Finally, the observation that the expansion of the universe is speeding up and not slowing down is counter to the assumption that the energy density of the universe is dominated by matter or radiation. This accelerated expansion can be explained by an additional component to the energy density of the universe commonly referred to as dark energy [317]. The dark energy component can be described with an unusual equation of state where the energy density is equal to the *negative* of its pressure, contrast this with the typical situations of radiation or baryonic matter which feature a positive equation of state.

One of the largest problems in the big bang cosmology that inflation solves is that of the horizon problem [317, 242, 243]. The cosmic microwave background was formed when the universe was roughly one million light years across while the fastest a particle can move is the speed of light. The CMB is observed to be homogenous across the whole sky yet no known particles could possibly have provided this level of uniformity through standard

local interactions. However, if the universe were much smaller at some point in time - small enough to have this degree of uniformity - and then rapidly expanded so that this uniformity was locked in then this would explain the homogeneity of the CMB. Additionally, this rapid expansion would drive the curvature of the universe, which could take several values, to that of the observed value of a flat universe. Much in the same way as an ant walking on a very large balloon would observe, the rapid expansion stretches out a patch of potentially curved spacetime to such a large extent that any observer would be unable to tell the difference between curved or flat. This expansion is naturally driven by including an additional scalar field which gives an energy density that alters the expansion rate - this is inflation. Others have proposed a similar mechanism, a new scalar field, which drives the much less rapid expansion associated with the dark energy - however, it should be noted that the dark energy could be accommodated by inclusion of a cosmological constant in Einstein's equation [237, 238, 239, 240, 317].

The main signals that the Standard Model are incomplete are the existence of dark matter and the nonzero value for neutrino masses. The issue of dark matter was discovered by studying the matter distribution of distant galaxies. A typical spiral galaxy can be treated as an extended body of mass that rotates around its center according to standard gravitational theory. Given the distribution of mass, which is obtained from the observable glowing stars, a prediction of the velocity as a function of distance - a galactic rotation curve - can be made and compared to an actual observed result. The stunning result was that the predicted rotation curve badly matched the actual rotation curve which implied that either gravity behaves differently at galactic scales or that there is additional matter to account for the deviation [317, 321, 322]. This debate was settled by the bullet cluster: a galactic cluster where two galaxies collided and a clear separation of a visible and dark components was observed through the use of gravitational lensing [19]. Additionally, the power spectrum of the CMB is sensitive to the amount of dark and regular matter - further confirming the problem of missing mass [63]. Nothing in the standard model of particle physics can produce the observed effects.

The nonzero masses of neutrinos also has its history rooted in astrophysics. The standard solar model is a model which describes the nuclear reactions and processes which power the

sun. The standard solar model predicts a certain flux of neutrinos being produced and delivered to the earth. Ray Davis had designed an experiment which would capture the total number of neutrinos from the sun by an ingenious use of using dry cleaning fluid as a reactant with the neutrinos. A large tank was filled with the dry cleaning fluid and as a neutrino interacted inside the tank, it would produce inert argon that would be detected after cleaning out the fluid. The striking result was that the predicted neutrino flux was only one third of that predicted by theory. Future experiments confirmed this phenomena and it was concluded that as neutrinos propagate, they change flavor. In other words, an electron neutrino may be produced at some source location but then as it travels to a detector location it may be observed as an electron, muon or tau neutrino [30, 31, 32]. The way to accommodate this observation in the Standard Model is to include non-zero mass and mixing terms for the neutrino sector.

The evidence for these missing pieces - neutrino masses, inflation, dark matter - is quite robust and represents some of the most important experimentally and observationally accessible problems which need to be addressed in high energy physics. Neutrino masses may augment the Standard Model most simply by including mass terms very similar to those in the quark sector. This generates an analog to the CKM matrix which has just as many parameters which need to be fit by experiment. The need to measure a host of parameters has generated an array of experiments which have lead to a measurement of all but only a few of these parameters [26, 28, 29, 33, 34, 35, 36, 37]. Inflation is described by the dynamics of a scalar field where a flat potential drives inflation until the field reaches its potential minimum - known as slow roll inflation. This scalar field induces quantum mechanical fluctuations which become causally disconnected from the universe, only to reenter at a later stage becoming the seeds of galactic structure growth [245, 241, 242, 243, 244]. This picture has been confirmed to unprecedented levels of accuracy by Planck and WMAP which fit the miniscule fluctuations in the CMB to the picture of inflation. Finally, the issue of dark matter has been observed in numerous galactic rotation curves, in the CMB and through gravitational lensing in the bullet cluster.

1.3 POSSIBLE SOLUTIONS AND FURTHER PROBLEMS: GUIDANCE FROM OBSERVATION AND EXPERIMENT

The standard slow roll inflation scenario is described by the dynamics of a scalar field where a flat potential drives inflation until the field reaches its potential minimum. This scalar field induces quantum mechanical fluctuations which become causally disconnected from the universe, only to reenter at a later stage becoming the seeds of galactic structure growth. During inflation the physical wavelengths of the fluctuations grow beyond the causal horizon, decoupling from the universe, only to re-enter in a later matter dominated period of expansion where the causal horizon grows faster than the scale factor [245, 241, 242, 243, 244]. This mismatch between the growth of scale factor and causal horizon results in the physical wavelengths re-entering the causal horizon - the previously disconnected physical wavelengths become smaller than the horizon and become causally correlated. The fluctuations in the scalar field translate into matter density fluctuations which, in turn, result in the production of tensor fluctuations. This arises naturally because matter is coupled to the gravitational field and fluctuations in the matter field will couple to produce gravitational fluctuations. This leads to a distinct prediction of inflation: the generation of scalar and tensor fluctuations [237, 238, 239, 240].

This picture of inflation has been confirmed to unprecedented levels of accuracy, by Planck and WMAP, but persistent and intriguing anomalies exist [249, 248, 63]. The most recent observations of the CMB [63] continue to find a statistically significant discrepancy between observation and the predictions of inflation where a power deficit $5-10\%$ at the largest scales is observed. This anomaly has been present in the previous COBE and WMAP [247, 249] missions and could potentially be explained through an alteration of the inflationary picture. This large scale power suppression has persisted in an otherwise consistent picture of ΛCDM and serves as motivation for deviations from the standard "slow roll" inflationary scenario. The possible origin of the large scale anomalies is far from settled and has generated much discussion as the observations at large scale reflect the most primordial signals that are within observational reach.

Observations of the CMB also provide a measurement on the number of neutrino species present in the early universe and this measurement seems to indicate that additional neutrino species may exist [61, 63]. The standard picture of ΛCDM cosmology describes the missing dark matter species as cold yet consideration of galactic scales suggests that a hotter, lighter dark matter candidate may be more successful [69, 70, 71, 72, 73, 74]. These cracks in both the standard cosmological and particle physics models have suggested that there is still more to be learned and discovered. Recent experiments in neutrino physics have even suggested that additional, more massive neutrinos which do not participate in Standard Model interactions may exist; these hypothetical objects are known as sterile neutrinos. [43, 44]

Neutrino masses and mixing are one of the clearest indicators that the standard model is incomplete. The standard model must be augmented to explain this phenomena and the possibility that neutrino masses are generated in the exact same manner as the other matter particles of the Standard Model remains possible. The most minimal model which includes the analog to the CKM matrix, known as the PMNS matrix, can describe the mixing of the neutrino sector but is not the most general way to include neutrino masses. Because neutrinos are neutral, it is possible for the neutrinos to be either Majorana or Dirac particles - Majorana particles and antiparticles are indistinguishable from one another. The minimal model which gives neutrino masses in the same manner as the rest of the Standard Model assumes that neutrinos are of a Dirac nature, but it is possible to include additional mass terms for neutrinos if they are Majorana particles. If neutrinos are Majorana particles, this opens up the possibility of multiple CP violating angles, lepton flavor violation and other potential observables [26, 28, 29, 31, 38].

If neutrinos are Majorana particles, there is the possibility of including an additional mass term which does not mix chirality - a mass term which is impossible for the other particle of the Standard Model. This is natural in see-saw models where this Majorana mass term is much larger than the Dirac mass term and, upon diagonalization of the mass matrix, this leads to a natural explanation of the lightness of the active type neutrinos [390]. This additional neutrino features no direct couplings to the standard model interactions and only interacts through mixing with the active types. This has earned the particle the moniker of

"sterile" neutrino and is one of the central themes of this thesis.

The theoretical motivations for sterile neutrinos stand on their own but are bolstered by anomalies in recent neutrino oscillation experiments. The oscillation experimental program has investigated the neutrino mixing parameter space which explains and measures the atmospheric and solar oscillation data [41, 42]. Older results from the LSND experiment [43] have suggested the possibility of the existence of a sterile neutrino with an eV mass - much larger than the active species. This experimental anomaly was also present in the MiniBooNE experiment [44], further bolstering the experimental hints. The presence of an electronvolt sterile neutrino is hotly debated and will likely not be resolved conclusively for several more years. Despite the controversy, these experimental anomalies and theoretical motivations serve as tantalizing motivation to consider the implications of sterile neutrinos in a cosmological and terrestrial context.

The extensions to the Standard Model which can potentially explain some or all of this phenomena are plentiful [318, 319] and the proliferation of models in the past few decades can be dizzying. The extensions range from extending spacetime into supersymmetry [20], additional spatial dimensions [21], Lorentz violations [22], technicolor [23], extensions to QCD [319], grand unified theories [24], and so on. With the perspective that there seems to be limitless parameter space to choose from, it seems natural to start at the most minimal model which is consistent with observations of particle physics and cosmology. The most minimal model available is known as the νMSM - a minimal model which provides masses for the neutrinos, a sterile neutrino dark matter candidate, enough CP violation to explain the matter-antimatter asymmetry and uses the Higgs boson as the inflaton [25]. This model augments the Standard Model with three right handed neutrino fields which could explain the short baseline anomalies, dark matter and provide a see-saw mechanism to explain the lightness of active neutrinos. In this thesis, we will frequently work under the assumption that the Standard Model is augmented in a similar way - with several right handed neutrinos.

The experimental confirmation of neutrino masses and mixing have ushered in a new era of exploration of physics beyond the Standard Model. Recent experiments suggest the existence of new species of neutrinos, sterile neutrinos, that do not feature standard model

interactions but mix with the active species [43, 44]. These sterile neutrinos may answer the dark matter question and the recent observations of 3.5 keV X-ray lines could herald the confirmation of this particle species as a dark matter component [351]. The problems associated with neutrinos may very well be entangled with the galactic scale problem in cosmology. This work focuses on topics in both cosmology and particle physics in the hopes of elucidating a better understanding of the early periods of inflation and how sterile neutrinos could solve the dark matter issue.

1.4 SUMMARY OF THE MAIN RESULTS

This thesis focuses on neutrino phenomenology associated with terrestrial experiments, inflationary cosmology, and sterile neutrinos as a dark matter candidate. This thesis attempts to bridge the realm of questions in particle physics and questions in cosmology while exploring timely issues on both fronts. This work is based on a series of articles and here we provide a brief summary with appropriate references.

Sterile Neutrino Phenomenology:

Based on the following papers:
(ref. [139]) L. Lello, D. Boyanovsky, Searching for sterile neutrinos from π/K decays, Phys. Rev. D 87, 073017 (2013)
(ref. [339]) L. Lello, D. Boyanovsky, Charged lepton mixing via heavy sterile neutrinos Nuclear Physics B, Volume 880, March 2014, Pages 109-133

In neutrino physics, part of this research focused on quantum entanglement and decoherence effects arising from the production of sterile neutrinos at current facilities. In chapter 2, several experimental signatures are proposed for identifying sterile neutrinos in accelerator based experiments as a result of the entanglement with other daughter particles in the production process. These ideas lead to a natural follow-up proposal which showed that heavy

sterile neutrinos would contribute to leptonic mixing and would lead to lepton flavor violation (such as $\mu \to e\gamma$) that could potentially be observed in current and proposed experiments - this is the subject of chapter 3. This work heavily borrowed techniques normally employed in quantum optics and applied the same ideas to questions relevant for terrestrial neutrino experiments to find novel potential signatures. These chapters were highly motivated by the short baseline anomalies of oscillation experiments [43, 44].

In chapter 2 we focus on the production of sterile neutrinos in current and proposed terrestrial experiments. The decay of a pion or kaon at rest can produce a monochromatic beam of charged leptons indicative of a massive sterile neutrino. The branching ratio and decay rate for this process is calculated. Additionally, in accelerator based neutrino oscillation experiments, the beam of neutrinos is produced through decaying pions travelling in a decay pipe. The decay process yields an entangled state between the daughter particles and the correlation is carried through to the propagating neutrino. We calculate the effect of a sterile neutrino on the appearance and disappearance signal for these oscillation experiments and show a nontrivial modification to the standard oscillation formula. The magnitude of these effects is shown for several current and proposed oscillation experiments.

Chapter 3 turns its attention from oscillation experiments to the phenomena of charged lepton mixing. The nonzero masses of neutrinos lead to a mismatch between the flavor and propagating mass eigenstates of the neutrinos. In the quark sector, a similar mismatch occurs but simply switching to the appropriate basis is possible because of the GIM mechanism - a reflection of unitarity of the CKM matrix. In the charged lepton sector, if we allow the possibility of sterile neutrino, a GIM type cancellation is not possible and leads to the possibility of mixing amongst the charged leptons. In this chapter, we compute the 1-loop corrections to the self energy of electrons and muons with a sterile neutrino in order to identify the correct mass eigenstates. This generates a flavor and mass mismatch that cannot be adjusted by switching bases as with the quark sector. From this, we calculate the contribution to current experiments that are searching for lepton flavor violating processes.

Quantum Correlations from Entanglement: Based on the following papers:

(ref. [198]) L. Lello, D. Boyanovsky, R. Holman, Entanglement entropy in particle decay, JHEP 2013 116

(ref. [299]) L. Lello, D. Boyanovsky, R. Holman, Superhorizon entanglement entropy from particle decay in inflation, JHEP 04 055 (2014)

(ref. [431]) D. Boyanovsky, L. Lello, Time evolution of cascade decay, New Journal of Physics 16 (2014) 063050

The production of two daughter particles which are correlated via quantum entanglement occurs from decay processes and, during inflation, can produce one daughter which is sub-horizon while the other is super-horizon. During an inflationary period, quantum fluctuations which are larger than the horizon will decouple from the system. This production process leads to a cross-horizon correlation which should persist throughout the cosmological evolution and have an imprint on the CMB. A similar phenomena occurs in standard experimental particle decays when one of the daughters remains unobserved. Part of this work studied these types of correlations in both settings, specifically those that could be relevant for terrestrial experiments in chapter 4 and those relevant for the inflationary paradigm in chapter 5. Additionally, with an eye towards similar phenomena resulting from decay chains, the preliminary work of chapter 6 has investigated similar issues relevant for cascade processes which could be applied towards further studies where quantum correlations could play a role.

The issue of entanglement entropy in standard particle decays is explored in chapter 4. Quantum entanglement, normally exploited in the realm of condensed matter physics, has been recently employed as a useful tool in particle physics experiments. Quantum correlations have been used in B meson experiments for studies on CP violation and the utility of entanglement for particle physics experiments is only just being recognized. We focus on a set of experiments where some of the particles produced are unobserved, leading to a loss of information. The natural example of this process is a particle decaying into a lepton-neutrino pair where the neutrino is unobserved. We calculate the loss of information, ie the entanglement entropy, of the system by computing the density matrix of the fully time evolved quantum state and then tracing over the unobserved particle. With this result, which

scales as the volume, we discuss and propose an experiment where this could be measured.

Changing to the realm of cosmology in chapter 5, we focus similar issues but in a de Sitter spacetime with an eye on observables in the CMB. In cosmology, the lack of energy conservation leads to the phenomena of a particle being able to decay into its own quanta. In a de Sitter quasi-inflationary spacetime, we investigate a scalar particle that decays into a pair of its own quanta, one of which is outside the causal horizon while the other lies within. The decay out of the causal horizon will be "unobserved" in the context of the observable universe and will act as a natural source of the loss of information. This decay into pairs and subsequent loss of information at the horizon is similar to the phenomena of Hawking radiation in black holes where particles pairs are produced on the edge of the event horizon. As with the case of a particle decay, we compute the fully time evolved density matrix and trace over the unobserved, superhorizon modes. This result scales as the volume as well, which in this case corresponds to the Hubble scale, and discuss the relation to non-gaussianties in the CMB.

In chapters 4,5 we had focused on results from a parent decaying into two daughters - both in Minkowski and de Sitter spacetimes - however, one can imagine a successive cascade scenario where one of the daughters decay further. In 6, we discuss the possibility of this in Minkowski spacetime as related to terrestrial experiments. We extend some of the results in the previous work to further the decay chain by one step, ie we allow for the two processes $A \to BC$, $B \to DE$. In the previous works, we obtained our results by using the fully time evolved quantum state resulting from a decay process to generate the significant results; however, the technical details the time evolved state involving a successive decay was unavailable. In this chapter, we focus on obtaining the time evolved state for this processes by treating the process as an initial value problem. The main result is obtained in another manner by using the Wigner-Weisskopf method and identifying the appropriate Markovian approximations. This result is discussed in the context of further cascades and the possibility of use in neutrino experiments searching for sterile neutrinos.

Initial Conditions in the Early Universe: Based on the following papers:

(ref. [311]) L. Lello, D. Boyanovsky, R. Holman, Pre-slow roll initial conditions: large scale power suppression and infrared aspects during inflation, Phys. Rev. D 89, 063533 (2014)
(ref. [283]) L. Lello, D. Boyanovsky, Tensor to scalar ratio and large scale power suppression from pre-slow roll initial conditions, JCAP 05 02 (2014)

Large scale anomalies in the temperature power spectrum of the CMB have persisted for over 20 years of observations and these anomalies may signal modifications to the standard cosmological model through the addition of a pre-inflationary epoch. We address how a very general pre-inflationary "fast-roll" stage would modify the dynamics of inflation in chapter 7 and the imprint that this stage could leave on the power spectrum of the CMB in chapter 8. In chapter 8, we show that a pre-inflationary fast-roll stage would leave a potentially observable signal in the tensor-to-scalar-ratio by imprinting oscillatory behavior that is not present in the standard inflationary paradigm. As mentioned previously, a generic prediction of inflation is the production of both scalar and tensor perturbations where the tensor-to-scalar ratio is the quantity which captures the relative magnitude between the two types of fluctuations. With future attempts to measure the primordial tensor fluctuations, or gravitational waves, the potential for observation of these oscillations in the near future is a possibility and could provide an observational window into the beginnings of inflation.

In chapter 7, we focus on the issues pertaining to how initial conditions can modify the infrared behavior of scalar fields in inflation - specifically the issue of dynamical mass generation and decay of particles. We consider non Bunch-Davies initial conditions which allow for the possibility of a fast roll stage prior to the slow roll inflationary stage. This fast roll stage enters through the equations of motion and leads to a modification of the Bogoliubov coefficients of the field quantization. We calculate the modification to the dynamical mass generation of a scalar field and the production rate for a particle to decay into a pair of subhorizon and superhorizon modes due to the presence of initial conditions. The fast roll initial conditions lead to a suppression in the CMB power spectra at large scales with a similar suppression reflected the mass generation, decay width and entanglement entropy.

Chapter 8 remains focused on the issue of fast roll generated non Bunch Davies initial

conditions but is more targeted towards the potential effect on CMB observables. The power spectra of curvature and tensor perturbations, ie primordial gravitational waves, are calculated in the context of modified fast roll initial conditions and, with these, the tensor to scalar ratio is computed. We calculate the modification to the low multipoles in the spectra and reproduce a suppression in the quadrupole - an anomaly that has persisted through the years and serves as a motivation for the initial condition modification. We show that these modified initial conditions lead to an oscillatory behavior in the tensor to scalar ratio which will hopefully become a possible observable signal and lead to another probe of the inflationary epoch.

New Sources of Sterile Neutrino Dark Matter: Based on the following papers:
(ref. [396]) L. Lello, D. Boyanovsky, Cosmological Implications of Light Sterile Neutrinos produced after the QCD Phase Transition Phys. Rev. D 91, 063502 (2015)
(ref. [447]) L. Lello, D. Boyanovsky, The case for mixed dark matter from sterile neutrinos JCAP 06 (2016) 011
L. Lello, D. Boyanovsky, R. Pisarski, Production of heavy sterile neutrinos from vector boson decay at $T \simeq M_W$, to be submitted

The dark matter problem is considered by investigating production mechanisms for a sterile neutrino dark matter candidate. Quantum kinetic theory is employed to study the production and "freeze out" of dark matter in the early universe; using these methods, general properties of the non-thermal sterile neutrino distribution function are studied - namely the approach to equilibrium and the dependence on the specific decay channel. Specifically, in chapters 9,10 the example of pion decay after the QCD phase transition is used to illustrate these effects and how this new non-thermal *light or heavy* sterile neutrino distribution could contribute to large scale structure. The example of pion decay is the mechanism used in terrestrial experiments which provide experimental hints of a sterile neutrino yet the cosmological relevance of this mechanism has gone largely ignored. In chapter 9, we show that this new dark matter candidate features a low momentum enhancement akin to the familiar resonant mixing scenario, which requires a lepton asymmetry, while the enhancement

we find is present even with vanishing chemical potential. In chapter 10, we show that this low momentum enhancement is more pronounced for heavier species which leads to a *colder* dark matter species, how to naturally produce a *mixed* dark matter scenario and, with this new distribution, the consequences for structure formation and contributions to CMB measurements are studied. In chapter 11, the production from other sources, such as the decay of heavy vector bosons at electroweak temperatures, is studied along with finite temperature corrections to the active-sterile mixing parameter.

The literature typically discusses a keV mass sterile neutrino dark matter candidate which is typically produced most prominently at temperatures near the QCD phase transition. In chapter 9, we focus on the production of a *light* sterile neutrino ($\lesssim 1MeV$) produced after the QCD phase transition from pion decays. The typical production mechanism for sterile neutrino dark matter is through mixing with the active species - production which is suppressed at higher temperatures - but production mechanisms outside of this standard scenario are not commonly discussed. If we take the QCD confinement temperature as the accepted region of production, then we should expect production through electroweak decays of strongly bound particles - the pion being the most abundant at these temperatures. We calculate the non equilibrium distribution function for light sterile neutrinos and compute the contribution to cosmological observables - dark matter abundance, free streaming length, phase space density. We show that this channel produces a non negligible fraction of a dark matter candidate.

Chapter 10 is an extension of the work in the previous chapter. We discuss the general production mechanisms available in the early universe that are usually not discussed in the literature. We identify several previously unstudied channels while discussing the general conditions for a sterile neutrino to come into thermodynamic equilibrium. We then return to the pion example and discuss the possibility of producing a heavier species from different channels. The heavier species of sterile neutrinos yields a colder distribution function than the light species, an observation which is a result of producing non-relativistic particles. The specific production channel of neutrinos yields different non-equilibrium distributions which yields a sort of kinematic entanglement. Due to the multiple channels available for a pion to decay ($\pi \to e\nu$,$\pi \to \mu\nu$), this naturally leads to a multicomponent dark matter - ie mixed

dark matter. This chapter illustrates that a non-equilibrium sterile neutrino dark matter candidate naturally leads to a scenario with multi component dark matter - a result which is amplified when considering the other production mechanisms.

In chapter 11, we study one of the production mechanisms which was mentioned in the prior chapter: decay of massive vector bosons. Being the carrier of the weak force, this production mechanism would be the natural source of sterile neutrinos if it were not for finite temperature suppression of active-sterile mixing. In the literature, the active-sterile mixing suppression at high temperature serves to dissuade consideration of this production mechanism but a precise evaluation is unavailable. We investigate how the mixing arises through proper identification of the propagating modes in the plasma - a technique which is not usually discussed in this context. We find that the active-sterile mixing is modified from the usual result of the literature by inclusion of the imaginary part of the self energy. With a more precise version of the active-sterile mixing, the production of sterile neutrinos from W, Z particles below the electroweak scale is investigated. Theoretical BBN calculations have yielded a stunning degree of agreement to observations of primordial elemental abundance with the notable exception of ^7Li where the calculations and observations feature a factor of 3 discrepancy. We find that this production mechanism can naturally lead to a nontrivial population of massive MeV sterile neutrinos which could be a solution to the ^7Li problem but that the contribution to a dark matter candidate directly would be negligible. It has been proposed that a heavy sterile neutrino can decay yielding a photon of the appropriate energy to react with primordial ^7Be in order to suppress the production of ^7Li. This mechanism would additionally yield a contribution to the relativistic degrees of freedom, N_{eff}, in the CMB which is within the bounds set by the Planck mission.

The remaining chapters expand on these points and provide the technical details involved in the calculations.

2.0 SEARCHING FOR STERILES FROM π/K DECAY.

Based on: (ref. [139])

L. Lello, D. Boyanovsky, Phys. Rev. D 87, 073017 (2013)

2.1 INTRODUCTION

Neutrino masses, mixing and oscillations are the clearest evidence yet of physics beyond the standard model [26, 27, 28, 29]. They provide an explanation of the solar neutrino problem [30, 31, 32] and have important phenomenological [26, 28, 29, 33, 34, 35, 36, 37], astrophysical [31, 38, 39] and cosmological [40] consequences. A remarkable series of experiments have confirmed mixing and oscillations among three "active" neutrinos with $\delta m^2 = 10^{-4}$–$10^{-3}\,\mathrm{eV}^2$ for atmospheric and solar oscillations respectively. The current bounds on these specifically are $\Delta m^2_{21} = 7.62 \times 10^{-5} eV^2$ (best fit) with a 1σ range $(7.43 - 7.81 \times 10^{-5} eV^2)$ and $\Delta m^2_{31} = 2.55 \times 10^{-3} eV^2$ (best fit) with a 1σ range $(2.46 - 2.61 \times 10^{-3} eV^2)$ respectively [41], for a complementary global analysis see[42].

However, several experimental hints have been accumulating that cannot be interpreted within the "standard paradigm" of mixing and oscillations among three "active" neutrinos with $\delta m^2 \simeq 10^{-4} - 10^{-3}$. Early results from the LSND experiment[43] have recently been confirmed by MiniBooNE running in antineutrino mode[44] both suggesting the possibility of new "sterile" neutrinos with $\delta m^2 \sim \mathrm{eV}^2$. The latest report from the MiniBooNE collaboration[45] on the combined $\nu_\mu \to \nu_e$ and $\overline{\nu}_\mu \to \overline{\nu}_e$ *appearance* data is consistent with neutrino oscillations with $0.01 < \Delta m^2 < 1.0\,\mathrm{eV}^2$. This is consistent with the evidence from

LSND antineutrino oscillations[43], which bolsters the case for the existence of sterile neutrinos; however, combined MiniBooNE/SciBooNE analysis[46] of the $\overline{\nu}_\mu$ *disappearance* data are consistent with *no short baseline disappearance* of $\overline{\nu}_\mu$. Recently, a re-examination of the antineutrino flux[47] in anticipation of the Double Chooz reactor experiment resulted in a small increase in the flux of about 3.5% for reactor experiments leading to a larger deficit of 5.7% suggesting a *reactor anomaly*[48]. If this deficit is the result of neutrino mixing and oscillation with baselines $L \lesssim 10 - 100\,\mathrm{m}$, it requires the existence of at least one sterile neutrino with $\delta m^2 \gtrsim 1.5\,\mathrm{eV}^2$ and mixing amplitude $\sin^2(2\theta) \simeq 0.115$[48]. Taken together these results may be explained by models that incorporate one or more sterile neutrinos that mix with the active ones[49, 50, 51, 52, 53, 54, 55, 56] including perhaps non-standard interactions[57]; although, there is some tension in the sterile neutrino interpretation of short-baseline anomalies[58]. These tensions present themselves in the "goodness of fit" parameter, which is obtained by comparing the fit of LSND with MiniBooNE antineutrino data and all other data, which is presently too low. A comprehensive review of short baseline oscillation experiments summarizes their interpretation in terms of one or more generations of sterile neutrinos[59, 60].

Hints for the existence of sterile neutrinos also emerge from cosmology. The analysis of the cosmic microwave background anisotropies by WMAP[61] suggests that the effective number of neutrino species is $N_{eff} = 3.84 \pm 0.40$ and $\sum(m_\nu) < 0.44\,eV$, suggesting the case for sterile neutrino(s) with $m \lesssim$ eV, however the recent results from (SPT), (ACT)[62] and PLANCK[63] weaken the bounds considerably. These bounds are obtained assuming 3 active neutrinos, 2 sterile neutrinos and incorporate CMB data, matter power spectrum information and a prior on the Hubble constant [64]. More recently stronger bounds on active-sterile neutrino mixing including Planck data has been reported[65]. Complementary cosmological data suggests that $N_{eff} > 3$ at the 95% confidence level[66]; although, accommodating an eV sterile neutrino requires a reassessment of other cosmological parameters[67]. For recent reviews on "light" sterile neutrinos see ref.[68]. Furthermore, sterile neutrinos with masses in the $\sim$ keV range *may* also be suitable warm dark matter candidates[69, 70, 71, 72, 73, 74] compatible with the ΛCDM model and may potentially solve the small scale problem. An experimental confirmation of sterile neutrinos would obviously bolster the argument for a

cosmologically relevant warm dark matter candidate.

When taken together, these emerging hints motivate several experimental proposals to search for sterile neutrinos (see the reviews in ref.[68]). Various experimental searches have been proposed, such as Higgs decay and matter interactions of relic sterile neutrinos[75], the end point of β-decay in ^{187}Re with a value of $Q = 2.5\,\mathrm{keV}$[76, 77] (although the statistics will be hindered by the long lifetime of the source $\simeq 4.3\times10^{10}$ years), and electron capture decays of $^{163}Ho \rightarrow ^{163}\ Dy$[78] with a Q-value $\simeq 2.2\,\mathrm{keV} - 2.8\,\mathrm{keV}$. More recently, the focus has turned on the possible new facilities at the "intensity frontier," one such proposal being project X at Fermilab[79] which would deliver high-power proton beams of energies ranging from 2.5-120 GeV and offers flexibility in the timing structure of beams. Another proposal involves using alternative high intensity sources[68, 80] such as mono-energetic electron neutrinos from an Ar^{37} source and detecting the nuclear recoil. There are also recent proposals to study sterile-active oscillations with pion and kaon decay at rest (DAR)[81, 82] where a cyclotron-based proton beam can be used to produce a low energy pion and muon decay-at-rest neutrino source as well as proposals that employee the use of muons from a storage ring[83]. In addition, the possibility of discrimination between heavy Dirac and Majorana sterile neutrinos[84] via $|\Delta L| = 2$ processes in high luminosity experiments[222] has been proposed, this is summarized in recent reviews[59, 60].

Goals: Our goals are the following:

- **a:)** Motivated by the possibility of high intensity sources, we assess the signals of *heavy* sterile neutrinos from meson (DAR) (both $\pi^-; K^-$) by focusing on searching for charged leptons of *negative helicity* (or positive helicity for their antiparticles in $\pi^+; K^+$ in (DAR)) in a setup akin to the Stern-Gerlach type experiment where opposite helicity components are spatially separated by a magnetic field gradient. Meson (DAR) produces a monochromatic beam of charged leptons back-to-back with (anti) neutrinos. Massive neutrinos yield a *negative helicity* component for the charged lepton which, in a collimated beam, may be separated from the (larger) positive helicity component by a magnetic field gradient. We study the branching ratio for the *negative helicity* component as a function of the sterile neutrino mass, as a complement to the search for monochromatic lines. We find

that for pion (DAR) the electron channel is the most efficient for $3\,\mathrm{MeV} \lesssim m_s \lesssim 135\,\mathrm{MeV}$ whereas for K-(DAR) both muon and electron channels are similar in the mass range allowed by the kinematics. We obtain an estimate for the upper bound on the branching ratio from previous experiments with typical upper bounds $Br \lesssim 10^{-8} - 10^{-6}$ perhaps accessible in the next generation of high intensity experiments.

- **b:)** We assess decoherence effects of sterile-active neutrino oscillations in *short baseline experiments* as a consequence of i) the decay width of the meson, and ii) the stopping distance of the charged lepton. As previously found in refs.[85, 86] the decay width of the meson leads to decoherence of oscillations quantified by the dimensionless ratio

$$\mathcal{R} = \frac{\delta m^2}{2E\Gamma_M}$$

where E is the neutrino energy and Γ_M is the meson decay width. For example, a Pion (DAR) with $E \simeq 30\,\mathrm{MeV}$ and $\Gamma_\pi \sim 2.5 \times 10^{-8}\,\mathrm{eV}$ leads to $\mathcal{R} \simeq \left(\delta m^2/\mathrm{eV}^2\right)$ and there could be considerable suppression of the appearance and disappearance probability in experiments with baseline $L \simeq 30 - 100\,\mathrm{mts}$[85, 86]. Another source of decoherence is the distance at which the charged lepton is stopped L_c: if the charged lepton is correlated with the emitted mass eigenstate over a long time scale, the quantum state is projected onto an energy eigenstate and oscillations are suppressed[85, 87]. Both effects, meson lifetime and charged lepton stopping scale, are sources of decoherence in sterile-active oscillations that are more prominent in short-baseline experiments and mass scales $\delta m^2 \simeq \mathrm{eV}^2$, as discussed in refs.[85, 86]. These effects can potentially impact the assessment of the sterile neutrino mass, mixing angle and CP-violation phases. We study both Dirac and Majorana neutrinos and show that these processes also affect CP-violating transitions. For Majorana neutrinos we study both $\Delta L = 0$ oscillations and $|\Delta L = 2|$ (L is lepton number) transitions from $\overline{\nu} \leftrightarrow \nu$ oscillations. We focus in detail on $3+2$ and $3+1$ schemes with new generations of sterile neutrinos and obtain the general CP-even and CP-odd expressions for the transition probabilities including $|\Delta L| = 2$ processes with Majorana neutrinos.

- **c:)** *If* sterile neutrinos are massive Majorana particles there are neutrino-antineutrino oscillations, these are lepton number violating transitions with $|\Delta L| = 2$. In short base-

line oscillation experiments, massive Majorana neutrinos yield two oscillation channels: the usual one with $\Delta L = 0$ and another with $|\Delta L| = 2$. While this latter channel is suppressed by the ratio m/E, we seek to study these lepton number violating oscillations in detail as potential discriminators between Dirac and Majorana neutrinos in future high luminosity experiments. Furthermore, neutrino-antineutrino oscillations can distinctly yield information about CP-violating Majorana phases[88] and one of our goals is to assess the impact of the above mentioned decoherence effects on the potential measurement of these transitions for new generations of sterile neutrinos.

Several appendices provide the technical details.

2.2 HEAVY STERILE NEUTRINOS IN RARE $\pi^{\pm}, K^{\pm}$ DECAYS AT REST:

In this work, our overarching goals are to assess the impact of sterile neutrinos in experimentally relevant situations. We begin this endeavor with the study of π/K decay at rest experiments and focus on helicity effects as potential experimental signals. The possibility of the existence of *heavy* "sterile" neutrino states had received early attention both theoretically[89] and experimentally[90, 91, 92, 93, 94, 95, 96, 97]; a review of the experimental bounds is presented in ref.[98]. In this section we analyze *possible* observational signatures of *heavy* sterile neutrinos in $\pi^-, K^- \to l^-_\alpha \, \overline{\nu}_\alpha$ decay at rest (DAR) but focus on *negative helicity charged leptons* (or positive helicity for π^+, K^+ decay). If the neutrino is massless, the charged lepton emerges from π, K (DAR) with right handed helicity (in the rest frame of the meson, which for (DAR) is the laboratory frame). However; if the neutrino is massive, a fraction of the charged lepton yield has left handed helicity. If the charged leptons are collimated along an axis z and there is a magnetic field that features a gradient along this direction, the situation is akin to the Stern-Gerlach experiment: the magnetic field gradient leads to a force $\vec{F} \propto -\vec{\nabla}(\vec{\mu} \cdot \vec{B})$ where $\vec{\mu}$ is the charged lepton magnetic moment. This force spatially separates the charged leptons with spins polarized parallel and antiparallel to the magnetic field gradient, just as in a Stern-Gerlach filter. The ratio of the helicity popula-

tion is determined by the branching ratio of the production process into *negative helicity charged lepton states*. Our goal is to obtain this branching ratio, which measures the relative intensity of the negative helicity states and *could* serve as a complement to the searches of monochromatic lines.

While there has been a substantial experimental effort[90, 91, 92, 93, 94, 95, 96, 97] searching for monochromatic lines associated with heavy sterile neutrinos from π, K decays, we are not aware of experimental efforts searching for *wrong* helicity charged lepton signals in mesons (DAR). The bounds obtained from the various experiments[90]-[97] are summarized as exclusion regions in ref. [98], which imply mixing angles (rather elements of the active-sterile mixing matrix) $\lesssim 10^{-6}$, making the branching ratios for these processes very small. However, high intensity beams as envisaged in the proposals[81, 82, 79, 68, 80] *may* provide the experimental setting to search for these signals complementing searches for monochromatic lines.

For a π or K meson, M, the interaction Hamiltonian for a $M \to l\,\bar{\nu}_l$ decay is given by

$$H_i = F_M \sum_{\alpha=e,\mu} \int d^3x \left[\,\overline{\Psi}_{l_\alpha}(\vec{x},t)\,\gamma^\mu \mathbb{L} \Psi_{\nu_\alpha}(\vec{x},t) J^M_\mu(\vec{x},t)\right] \quad ; \quad \mathbb{L} = \frac{1}{2}(1-\gamma^5) \tag{2.2.1}$$

where the label α refers to the charged leptons, $J^M_\mu(\vec{x},t) = i\partial_\mu M(\vec{x},t)$ and M is a complex (interpolating) field that describes the charged pseudoscalar mesons $M = \pi^-, K^-$. For a π^- meson, we have that $F_\pi = \sqrt{2}\,G_F\,V_{ud}\,f_\pi$ and for the $K^\pm$ meson, we have that $F_K = \sqrt{2}\,G_F V_{us}\,f_K$, where $f_{\pi,K}$ are the decay constants. The flavor neutrino fields and the fields that create/annihilate neutrino mass eigenstates are related by

$$\Psi_{\nu_\alpha} = \sum_j U_{\alpha j} \Psi_{\nu_j}\,. \tag{2.2.2}$$

For n generations of Dirac neutrinos the matrix U is $n \times n$, unitary and features $(n-1)(n-2)/2$ CP-violating Dirac phases. For Majorana neutrinos

$$U \to \widetilde{U} = U\,D \quad ; \quad D = \mathrm{diag}\left[e^{i\theta_1/2}, e^{i\theta_2/2}, \cdots, e^{i\theta_n/2}\right] \tag{2.2.3}$$

where U is the mixing matrix for Dirac neutrinos and we have allowed an inconsequential overall phase. It follows that

$$\widetilde{U}_{\alpha j} = U_{\alpha j}\, e^{i\theta_j/2}\,. \tag{2.2.4}$$

The Majorana CP-violating phases, $\theta_i - \theta_j$, *only* contribute to $\nu \leftrightarrow \overline{\nu}$ oscillations and $|\Delta L| = 2$ processes[88] which will be studied in detail in section (2.5).

The details of the quantization of the different fields are provided in appendix (2.8.1). From these results, it follows that, after integration over the spatial variables, the relevant Hamiltonian to obtain the production amplitudes is given by (see appendix (2.8.1) for notation)

$$H_i = \frac{F_M}{\sqrt{V}} \sum_{\vec{q},\vec{p}} \sum_{h,h'} \sum_{\alpha,j} U_{\alpha j} \frac{\left[\overline{\psi}_{l_\alpha}(\vec{k},h)\gamma^\mu \mathbb{L} \psi_{\nu_j}(\vec{q},h') p_\mu (M^+_{\vec{p}} - M^-_{\vec{p}})\right]}{\sqrt{8E_M(p)E_\alpha(k)E_j(q)}} \quad ; \quad \vec{k} = \vec{p} + \vec{q} \tag{2.2.5}$$

where the Fermi quantum fields, ψ_{ν_j}, are expanded as in (2.8.1) either for Dirac or Majorana fermions.

We identify the production matrix element $M^-(\vec{p}) \to l_\alpha(\vec{k})\, \overline{\nu}_\alpha(\vec{q})$ as

$$\mathcal{M}^{\mathcal{P}}_{\alpha,\alpha}(\vec{k},\vec{q},h,h') = \sum_j U_{\alpha,j}\ \mathcal{M}^{\mathcal{P}}_{\alpha j}(\vec{k},\vec{q},h,h') \tag{2.2.6}$$

where

$$\mathcal{M}^{\mathcal{P}}_{\alpha,j}(\vec{k},\vec{q},h,h') = F_M\, \overline{\mathcal{U}}_{\alpha,h}(\vec{k})\, \gamma^\mu \mathbb{L} \mathcal{V}_{j,h'}(\vec{q})\, p_\mu \quad ; \quad \vec{k} = \vec{p} + \vec{q} \tag{2.2.7}$$

is the transition matrix element for meson decay into a charged lepton, α, and an antineutrino mass eigenstate, j. For Dirac neutrinos, the spinors $\mathcal{V}_{j,h'}(\vec{q})$ in (2.2.7) are given by (2.8.8), whereas for Majorana neutrinos

$$\mathcal{V}_{j,h'}(\vec{q}) \to \mathcal{U}^c_{j,h'}(-\vec{q}) \tag{2.2.8}$$

given by (2.8.18) and the mixing matrix $U \to \widetilde{U}$ given by (2.2.3,2.2.4). The separation of helicity contributions is frame dependent and the most clear identification of processes that reveals a massive neutrino is provided by the decay of the pseudoscalar meson at rest ($\vec{p} = 0$) so that the laboratory coincides with the rest frame of the meson and helicity states

are unambiguously recognized. The contributions to the production amplitude from the different helicity states in (DAR) are given by

$$\begin{aligned}
&\overline{\mathcal{U}}_{\alpha,+}(\vec{q})\,\gamma^{\mu}\mathbb{L}\mathcal{V}_{j,+}(\vec{q})\,p_{\mu} = -m_M\varepsilon_l N_l N_{\bar{\nu}} \\
&\overline{\mathcal{U}}_{\alpha,+}(\vec{q})\,\gamma^{\mu}\mathbb{L}\mathcal{V}_{j,-}(\vec{q})\,p_{\mu} = 0 \\
&\overline{\mathcal{U}}_{\alpha,-}(\vec{q})\,\gamma^{\mu}\mathbb{L}\mathcal{V}_{j,+}(\vec{q})\,p_{\mu} = 0 \\
&\overline{\mathcal{U}}_{\alpha,-}(\vec{q})\,\gamma^{\mu}\mathbb{L}\mathcal{V}_{j,-}(\vec{q})\,p_{\mu} = m_M\varepsilon_{\bar{\nu}} N_l N_{\bar{\nu}}
\end{aligned} \tag{2.2.9}$$

where (see notation in appendix (2.8.1))

$$\varepsilon_a(q) = \frac{m_a}{E_a(q)+q} \;\;;\;\; N_a = \sqrt{E_a(q)+q} \;\;;\;\; E_a(q) = \sqrt{q^2+m_a^2} \;\;;\;\; a = l,\nu \tag{2.2.10}$$

Gathering these results, we obtain the helicity contributions to the π, K decay widths either for Dirac or Majorana neutrinos:

$$\Gamma^{++}_{\pi/K\to l\bar{\nu}_s} = \frac{G_F^2}{4\pi}|U_{ls}|^2\,|V_{ud/us}|^2\,f^2_{\pi/K}\;q^*\,m_l^2\left[\frac{E_{\nu_s}(q^*)+q^*}{E_l(q^*)+q^*}\right] \tag{2.2.11}$$

$$\Gamma^{--}_{\pi/K\to l\bar{\nu}_s} = \frac{G_F^2}{4\pi}|U_{ls}|^2\,|V_{ud/us}|^2\,f^2_{\pi/K}\;q^*\,m_{\nu_s}^2\left[\frac{E_l(q^*)+q^*}{E_{\nu_s}(q^*)+q^*}\right] \tag{2.2.12}$$

where

$$q^* = \frac{1}{2m_M}\left[\left(m_M^2-(m_l+m_s)^2\right)\left(m_M^2-(m_l-m_s)^2\right)\right]^{\frac{1}{2}} \;\;;\;\; m_s \leq m_M - m_l \tag{2.2.13}$$

and here we refer to the heavy sterile *mass eigenstate* as $\underline{s}$ rather than identifying it with a fourth or fifth generation.

In the limit $m_s \to 0$, the usual result for π, K decay at rest, where the antineutrino and the lepton are both right handed polarized, is obtained. We are particularly interested in the branching ratio for the process in which both the antineutrino and the charged lepton feature left handed helicity, given by (2.2.12). The branching ratio for this process is obtained by normalizing to the total meson width and since these are rare processes, we can instead normalize to the proxy to the total width

$$\Gamma^{tot}_{\pi/K} \equiv \frac{\Gamma_{\pi/K\to\mu\bar{\nu}}}{Br(\pi/K\to\mu\bar{\nu})} \tag{2.2.14}$$

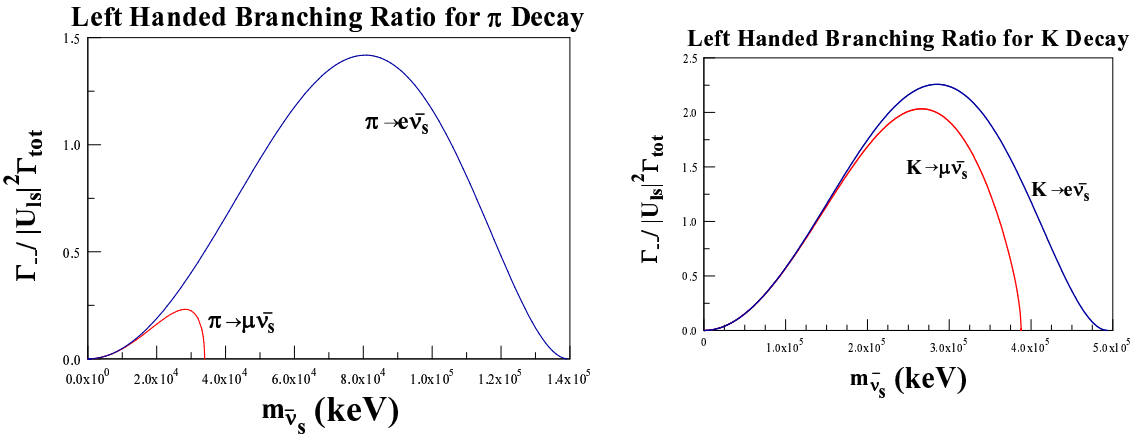


Figure 1: Left panel:$Br^{--}_{\pi\to\mu,e\bar{\nu_s}}/|U_{ls}|^2$, right panel:$Br^{--}_{K\to\mu,e\bar{\nu_s}}/|U_{ls}|^2$ vs. m_s for $l=\mu,e$.

where $Br(\pi/K\to\mu\bar{\nu})=0.999,0.635$ is the branching ratio for the purely leptonic decay into muons and *massless* neutrinos for π,K decay respectively. Specifically, we have

$$Br^{--}_{M\to l\bar{\nu_s}}\equiv\frac{\Gamma^{--}_{M\to l\bar{\nu_s}}}{\Gamma^{tot}_{M}}=|U_{ls}|^2\,\frac{2\,Br(M\to\mu\bar{\nu})\;q^*\,m^2_{\nu_s}}{m^2_\mu m_M\left(1-\frac{m^2_\mu}{m^2_M}\right)^2}\left[\frac{E_l(q^*)+q^*}{E_{\nu_s}(q^*)+q^*}\right] \tag{2.2.15}$$

Fig.(1) show the branching ratios (2.2.15) for $\pi\to\mu,e\,\bar{\nu}_s$ and $K\to\mu,e\,\bar{\nu}_s$ respectively. For π (DAR), the *electron* channel offers a larger window simply because of the larger amount of phase space available whereas the maximum mass available for a heavy sterile neutrino in the muon channel is $\sim 33.92\,\mathrm{MeV}$.

A Stern-Gerlach experiment:

In meson (DAR), the presence of a heavy sterile neutrino is manifest as a monochromatic line in the charged lepton spectrum at kinetic energy

$$T_l(q^*)=\frac{1}{2m_M}\Big[(m_M-m_l)^2-m_s^2\Big]\quad;\quad m_s\leq(m_M-m_l)\,. \tag{2.2.16}$$

The negative helicity component of the charged lepton in (DAR) provides another manifestation of a massive sterile neutrino which can be exploited in an experiment to complement

the search of monochromatic peaks in the charged lepton spectrum. The experimental setup to exploit the negative helicity component (or positive helicity for the opposite charged meson and charged lepton) should be akin to the original Stern-Gerlach experiment to separate spin components. In this case, the relevant quantity is helicity; therefore, consider collimating the charged leptons in (DAR) along a $z-axis$ and setting up a magnetic field with *a gradient along this direction* so that the direction of motion of the collimated charged leptons coincide with the direction of the gradient of the magnetic field. Under these circumstances, there is a magnetic force acting on the charged leptons

$$F_z \propto -h\frac{dB_z}{dz}\,, \tag{2.2.17}$$

where h is the helicity component; thus, opposite helicity components *separate spatially* and the fraction of negative helicity charged leptons is measured by the branching ratio (2.2.15). Therefore, searching for spatially separated domains of charged leptons *in combination* with a monochromatic line, *may* provide a more robust signature of heavy sterile neutrinos and allow extraction of the mixing matrix element $|U_{ls}|$: the mass of the sterile neutrino is inferred from the peak in the monochromatic spectrum while the ratio of abundances of the helicity components is determined by the branching ratio (2.2.15); therefore, with the input for q^* obtained from the peak in the monochromatic line and the measurement of the ratio of abundances of helicity states, the branching ratio (2.2.15) yields $|U_{ls}|$.

An estimate for the upper bound on the branching ratios, Br^{--}, given by (2.2.15) can be obtained from the summary of the bounds on the mixing matrix elements, $|U_{ls}|^2$, provided in ref.[98] for $l=\mu$: the exclusion region for π (DAR) from the μ spectrum yields an upper bound

$$|U_{\mu s}|^2 \lesssim 10^{-5} \;\;;\;\; 3\,\text{MeV} \lesssim m_s \lesssim 33\,\text{MeV} \tag{2.2.18}$$

and for K (DAR)

$$|U_{\mu s}|^2 \lesssim 10^{-6} - 10^{-5} \;\;;\;\; 30\,\text{MeV} \lesssim m_s \lesssim 330\,\text{MeV}\,. \tag{2.2.19}$$

The experiments[90, 91, 92, 93, 94, 95, 96, 97] on which the bounds in ref.[98] are based, search for monochromatic peaks in the muon spectrum, both from π, K (DAR). Ref.[99]

reported an upper limit $|U_{es}|^2 < 10^{-7}$ $(90\% C.L.)$ for $30 < m_s < 130\,\mathrm{MeV}$, therefore we find from fig. (1) that the upper bound for $Br^{--}_{\pi\to\mu,e\bar{\nu_s}}$

$$Br^{--}_{\pi\to\mu\,\bar{\nu_s}} \lesssim 10^{-8} - 10^{-7} \;\; ; \;\; 3\,\mathrm{MeV} \lesssim m_s \lesssim 33\,\mathrm{MeV} \tag{2.2.20}$$

$$Br^{--}_{\pi\to e\,\bar{\nu_s}} \lesssim 10^{-9} - 10^{-7} \;\; ; \;\; 30\,\mathrm{MeV} \lesssim m_s \lesssim 130\,\mathrm{MeV} \tag{2.2.21}$$

The small m_s region is obviously suppressed by the m_s^2/m_l^2 factor whereas the region near the kinematic edge is suppressed by phase space. For π decay, the electron channel is the most favorable to study the intermediate mass region $\simeq 3\,\mathrm{MeV} \lesssim m_s \lesssim 135\,\mathrm{MeV}$ with typical upper bounds on the branching ratios $10^{-8} - 10^{-6}$.

For K decay, both μ, e channels yield similar branching ratios with upper bounds in the range

$$Br^{--}_{K\to\mu,e\,\bar{\nu_s}} \lesssim 10^{-9} - 10^{-6} \text{ for } \begin{cases} 4\,\mathrm{MeV} \lesssim m_s \lesssim 360\,\mathrm{MeV}\;\; (\mu - \text{channel}) \\ 4\,\mathrm{MeV} \lesssim m_s \lesssim 414\,\mathrm{MeV}\;\; (e - \text{channel}) \end{cases} . \tag{2.2.22}$$

The "low" mass region of cosmological interest, $m_s \simeq$ few keV, is much more challenging. The experimental results of refs.[90, 91, 92, 93, 94, 95, 96, 97] and the analysis of ref.[98] do not provide reliable upper bounds; however, bounds for this mass range emerge from cosmology: a "heavy" sterile neutrino can decay into a photon and a light active neutrino, which, for $m_s \simeq$ keV, leads to an X-ray line. Cosmological constraints are summarized in the review articles in refs.[71, 72] with an upper bound $|U_{ls}|^2 \simeq 10^{-10} - 10^{-9}$ which would make the branching ratios exceedingly small, even for the high intensity sources envisaged.

To the best of our knowledge, Shrock[89][1] provided an early proposal to use *polarization* in combination with monochromatic line searches to obtain an assessment of neutrino masses and mixing. Our study differs from this earlier study in two main aspects: i) we advocate using combinations of magnetic fields in a Stern-Gerlach-type setup to *separate* the different helicity components. The relative abundance of the "wrong" helicity is determined by the branching ratio obtained above. This is important, while the polarization will be dominated by the lighter active-like neutrinos because they mix with larger mixing angles the *separation* of helicity components by magnetic fields, if experimentally feasible, could result in a clearer

[1] We thank R. Shrock for making us aware of his early work on these aspects.

signal. ii) Separating the helicity components via magnetic field configurations does not require searching for monochromatic lines and is an independent and *complementary* method. The proposal of ref.[89] requires first identifying the monochromatic lines and after this identification measuring the polarization, both aspects must be combined in this proposal to extract information perhaps increasing the challenge from the observational perspective.

A firmer assessment of whether the Stern-Gerlach type experiments, combined with searches of monochromatic peaks in π, K (DAR), are feasible in determining the masses of "heavy" sterile neutrinos calls for a detailed understanding of backgrounds which is a task that is beyond the scope of this article. Furthermore the above results only apply for $V-A$ weak interactions, therefore if sterile neutrinos feature non-standard weak interactions a re-assessment of the results is required[89].

2.3 OSCILLATIONS IN SHORT-BASELINE EXPERIMENTS

For short baseline oscillation experiments, the relevant range of neutrino mass differences is $\delta m^2 \simeq (\mathrm{eV})^2$. A detailed analysis of oscillation phenomena requires an understanding of the production and detection process. In ref.[85] a quantum field theoretical generalization of the Wigner-Weisskopf method[100, 101] was introduced to obtain the correct quantum state arising from the decay of the parent particle. A previous treatment of the correlations of the decay product within a Wigner-Weisskopf approach to semiclassical wave packets was originally studied in ref.[105] and the dynamics of propagation were studied in ref.[106] in simple models. In ref.[85], the method was implemented in a simple quantum field theory model of charged current interactions and several aspects were found to be much more general, such as the decoherence effects associated with the lifetime of the decaying parent particle as well as the observation (or stopping) of the charged lepton produced as partner of the neutrino in a charged current interaction vertex.

Many of these aspects were found also in refs.[86] in a different formulation but without explicitly obtaining the quantum mechanical state that describes the decay products.

Meson decay leads to a correlated state of the charged lepton and the neutrino, a quantum entangled state[87, 85, 102], the entanglement being a consequence of the kinematics and conservation laws pertinent to the decay[103]. As originally observed in ref.[87] and analyzed in detail in refs.[85, 102], quantum entanglement leads to decoherence in neutrino oscillations which is a result that has been confirmed more recently in [86, 104] within a different approach.

In this article, we generalize the quantum field theoretical Wigner-Weisskopf method introduced in ref.[85] to describe (pseudoscalar) meson decay via charged current interactions *in the standard model*, including all aspects of the interactions both for *Dirac and Majorana neutrinos*. An alternative formulation is offered in ref.[86]; however, the full quantum field theoretical Wigner-Weisskopf method not only illuminates clearly the *quantum entanglement and correlations* between the charged lepton and neutrino states both in momentum and helicity, but also allows a systematic study of Dirac and Majorana fermions including the dynamics of $\nu \leftrightarrow \overline{\nu}$ oscillations and $|\Delta L| = 2$ processes discussed in detail in section (2.5).

2.3.1 Production from meson decay:

In appendix 2.8.2 (see also ref.[175] for more details), we implement a quantum field theoretical version of Wigner-Weisskopf theory and we find the Schroedinger picture quantum state that results from pseudoscalar meson decay which is given by

$$\begin{aligned} |M^-_{\vec{p}}(t))\rangle \;&=\; e^{-iE_M(p)\,t}\, e^{-\Gamma_M(p)\,\frac{t}{2}}\, |M^-_{\vec{p}}(0)\rangle - \sum_{\vec{q},\alpha j,h,h'} \Big\{ U_{\alpha j}\, \Pi^{\mathcal{P}}_{\alpha j}\, \mathcal{M}^{\mathcal{P}}_{\alpha j}(\vec{k},\vec{q},h,h')\, \mathcal{F}_{\alpha j}[\vec{k},\vec{q};t] \\ &\times\; e^{-i(E_\alpha(k)+E_j(q))t}\, |l^-_\alpha(h,\vec{k})\rangle\, |\overline{\nu}_j(h',-\vec{q})\rangle \Big\} \;\;;\;\; \vec{k} = \vec{p}+\vec{q}, \end{aligned} \tag{2.3.1}$$

where

$$\Pi^{\mathcal{P}}_{\alpha j} = \frac{1}{[8V E_M(p) E_\alpha(k) E_j(q)]^{\frac{1}{2}}}\,. \tag{2.3.2}$$

Although we consider plane wave states, the generalization to wave-packets is straightforward and we comment on the wave-packet approach in section (2.6.2). The production

matrix element $\mathcal{M}^{\mathcal{P}}_{\alpha j}(\vec{k},\vec{q},h,h')$ is given by (2.2.7), (see eqn. (2.8.45) in appendix (2.8.2)) $\Gamma_M(p) = m_M\Gamma_M/E_M(p)$ where Γ_M is the decay width in the rest frame of the meson, and

$$\mathcal{F}_{\alpha j}[\vec{k},\vec{q};t] = \frac{1-e^{-i\Big(E_M(p)-E_\alpha(k)-E_j(q)\Big)t}\, e^{-\Gamma_M(p)\frac{t}{2}}}{E_M(p)-E_\alpha(k)-E_j(q)-i\frac{\Gamma_M(p)}{2}}\,. \tag{2.3.3}$$

The second term in (2.3.1) reveals that the emerging charged lepton and neutrino are entangled both in momentum and in helicity.

The factor $\mathcal{F}_{\alpha j}[\vec{k},\vec{q};t]$ encodes the time dependence of the production process. In order to understand the content of this factor, consider the case $\Gamma_M = 0$. In this case,

$$\mathcal{F}_{\alpha j}[\vec{k},\vec{q};t] = e^{-i(E_M-E_\alpha-E_j)\frac{t}{2}}\,\frac{2i\,\sin\Big[(E_M-E_\alpha-E_j)\frac{t}{2}\Big]}{\Big[E_M-E_\alpha-E_j\Big]} \overset{t\to\infty}{\to} 2\pi i\delta(E_M-E_\alpha-E_j) \tag{2.3.4}$$

namely, in the long time limit, this function describes *energy conservation at the production vertex*. The width of the decaying meson state determines a time (or energy) uncertainty and, either for a narrow width or large time, the function $\mathcal{F}_{\alpha j}[\vec{k},\vec{q};t]$ is strongly peaked at $E_\alpha + E_j \simeq E_M$ which describes *approximate energy conservation* within the time or width uncertainty.

In a typical experiment, the charged lepton produced by pion (kaon) decay is stopped shortly after the end of the pion decay pipe, at which point the correlated quantum state after the neutrino state is *disentangled* by the observation, capture or absorption of the charged lepton at t_c.

If the charged lepton l_α is observed, or absorbed with momentum $\vec{k}$ and helicity projection h_i at time t_c, the wave function is projected onto the state $\langle l^-_\alpha(h_i,\vec{k})|$ and the correct (anti) neutrino state that propagates is given by

$$|\widetilde{\overline{\nu}}(\vec{q};h_i)\rangle = -e^{-iE_\alpha(k)t_c}\sum_{j,h'} U_{\alpha j}\,\Pi^{\mathcal{P}}_{\alpha j}\,\mathcal{M}^{\mathcal{P}}_{\alpha j}(\vec{k},\vec{q},h_i,h')\,\mathcal{F}_{\alpha j}[\vec{k},\vec{q};t_c]\,e^{-iE_j(q)t_c}\,|\overline{\nu}_j(h',-\vec{q})\rangle\,, \tag{2.3.5}$$

where $\vec{q}=\vec{p}-\vec{k}$. This neutrino state still carries the label h_i as a consequence of the helicity entanglement with the measured charged lepton.

We note that *if* $\mathcal{M}^{\mathcal{P}}_{\alpha j}, \mathcal{F}_{\alpha j}, E_j$ are all independent of the mass of the neutrino, j, these factors can be taken out of the sum and the resulting (anti) neutrino state is proportional to the familiar Pontecorvo coherent superposition of mass eigenstates. We will analyze this approximation below after assessing the total transition amplitude from production to detection; however, before doing so, it proves illuminating to understand the *normalization* of the state (2.3.5).

$$\mathcal{N}_\nu(\vec{q}; h_i) \equiv \langle \widetilde{\overline{\nu}}(\vec{q}; h_i) | \widetilde{\overline{\nu}}(\vec{q}; h_i) \rangle = \sum_{j,h'} \left| U_{\alpha j} \right|^2 \left| \Pi^{\mathcal{P}}_{\alpha j} \, \mathcal{M}^{\mathcal{P}}_{\alpha j}(\vec{k}, \vec{q}, h_i, h') \right|^2 \left| \mathcal{F}_{\alpha j}[\vec{k}, \vec{q}; t_c] \right|^2 . \quad (2.3.6)$$

In the narrow width limit, the function $\left| \mathcal{F}_{\alpha j}[\vec{k}, \vec{q}; t_c] \right|^2$ becomes $\propto \delta\big(E_M(p) - E_\alpha(k) - E_j(q)\big)$ and the proportionality constant can be obtained by integrating this function in the variable $\mathcal{E} = E_M(p) - E_\alpha(k) - E_j(q)$, from which we find

$$\left| \mathcal{F}_{\alpha j}[\vec{k}, \vec{q}; t_c] \right|^2 = \frac{2\pi}{\Gamma_M(p)} \left[1 - e^{-\Gamma_M(p) t_c} \right] \delta\big(E_M(p) - E_\alpha(k) - E_j(q)\big) . \quad (2.3.7)$$

Therefore

$$\mathcal{N}_\nu(\vec{q}; h_i) = \frac{\left[1 - e^{-\Gamma_M(p) t_c} \right]}{\Gamma_M(p)} \sum_{j,h'} \left| U_{\alpha j} \right|^2 \left| \Pi^{\mathcal{P}}_{\alpha j} \, \mathcal{M}^{\mathcal{P}}_{\alpha j}(\vec{k}, \vec{q}, h_i, h') \right|^2 2\pi \, \delta\big(E_M(p) - E_\alpha(k) - E_j(q)\big) . \quad (2.3.8)$$

In appendix (2.8.3), we obtain the relation between the normalization (2.3.8), the partial and total decay width of the meson and the number density of charged leptons produced by meson decay during a time t_c. While ref.[38] discusses the normalization of the neutrino state[2], to the best of our knowledge, the relation of the neutrino normalization to the number density of charged leptons produced has not been recognized previously.

[2] See section (8.1.1), pages 285,286 in ref.[38].

2.3.2 Detection via a charged current vertex:

In what follows we assume the neutrino to be described by a Dirac fermion, extending the discussion to Majorana fermions in section (2.5). We note here that the Dirac or Majorana nature is irrelevant for the $\Delta L = 0$ process considered here but plays a nontrivial role in section (2.5).

Consider the case in which the (anti)neutrino is detected via a charged current event $\overline{\nu}\, N \to l^+_\beta\, N'$ at a detector situated at a baseline L (fig.(2)).

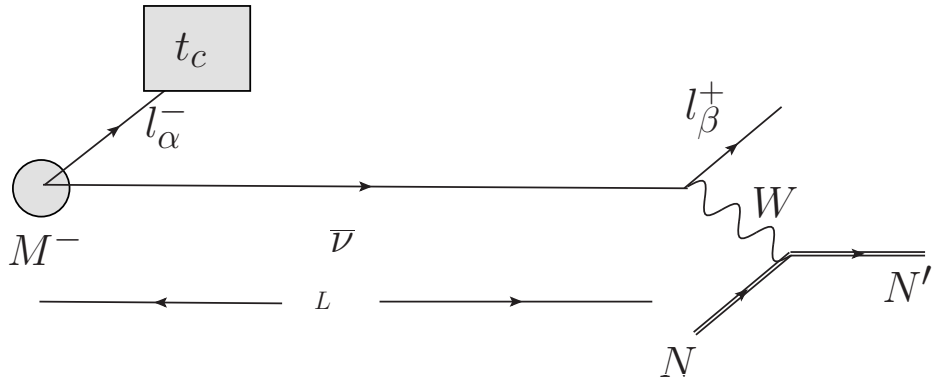


Figure 2: Production via $M^- \to l^-_\alpha \overline{\nu}$ detection via a charged current vertex $\overline{\nu}\, N \to l^+_\beta\, N'$ at a baseline L with N, N' nucleons or nuclear targets. The charged lepton l^-_α produced with the antineutrino is observed, absorbed or decays at a time t_c.

The Schroedinger picture quantum states that describe the initial and final states are

$$|i\rangle = |\widetilde{\overline{\nu}}; N\rangle = |\widetilde{\overline{\nu}}\rangle \otimes |N\rangle_{\mathcal{D}} \quad ; \quad |f\rangle = |l^+_\beta; N'\rangle = |l^+_\beta\rangle \otimes |N'\rangle \tag{2.3.9}$$

where $|\widetilde{\overline{\nu}}\rangle$ is given by (2.3.5), the state $|N\rangle_{\mathcal{D}}$ describes a nucleon or nuclear target localized at the detector and the outgoing charged lepton is measured with helicity h_f. The transition amplitude in the Schroedinger picture is given by

$$\mathcal{T}_{i\to f} = \langle f|e^{-iH(t_D - t_c)}|i\rangle \simeq -ie^{-iE_F t_D} \int_{t_c}^{t_D} e^{iE_F t'} \langle f|H_i\, e^{-iH_0 t'}\, e^{iH_0 t_c}|i\rangle\, dt' \tag{2.3.10}$$

where $E_F = E_{l_\beta} + E_{N'}$ is the total energy of the final state, and H_0, H_i, H are the unperturbed, interaction and total Hamiltonians respectively. To obtain this expression we have used $e^{-iH(t_D - t_c)} = e^{-iH_0 t_D}\, U(t_D, t_c)\, e^{iH_0 t_c}$ and $U(t_D, t_c)$ is the usual time evolution operator in the interaction picture.

Up to an irrelevant overall phase we find

$$\mathcal{T}_{i\to f} = \sum_{j,h'} \Pi^{\mathcal{P}}_{\alpha j} U_{\alpha j}\, \mathcal{M}^{\mathcal{P}}_{\alpha j}\, \mathcal{F}_{\alpha j}\, G_{\beta j}\, \langle l^+_\beta\ ; N'|H_i|\overline{\nu}_{j,h'}\ ; N\rangle \tag{2.3.11}$$

where $\Pi^{\mathcal{P}}_{\alpha j}$ is given by eqn. (2.3.2), we have suppressed the indices to avoid cluttering the notation and introduced

$$G_{\beta j} = e^{\frac{i}{2}(E_F - E_N - E_j)(t_D + t_c)}\ \frac{2\,\sin\left[\frac{1}{2}(E_F - E_N - E_j)(t_D - t_c)\right]}{\Big[E_F - E_N - E_j\Big]}\,. \tag{2.3.12}$$

The relevant interaction Hamiltonian is given by

$$H_i = U^*_{\beta j}\ \sqrt{2} G_F \int \overline{\Psi}_{\nu_j}(\vec{x}) \gamma^\mu \mathbb{L}\, \Psi_{l_\beta}(\vec{x})\, \mathcal{J}^{(N,N')}_\mu(\vec{x}) d^3x + h.c. \tag{2.3.13}$$

where $\mathcal{J}^{(N,N')}_\mu(\vec{x})$ is the hadron current with matrix element[3]

$$\langle N|\mathcal{J}^{(N,N')}_\mu(\vec{x})|N'\rangle \equiv \sum_P \frac{j^{N,N'}_\mu(P)}{\sqrt{4V E_N E_{N'}}}\, e^{i\vec{P}\cdot\vec{x}}\,. \tag{2.3.14}$$

leading to the matrix element

$$\langle l^+_\beta\ ; N'|H_{int}|\overline{\nu}_j\ ; N\rangle = U^*_{\beta j}\, \Pi^{\mathcal{D}}_{\beta j}\, \mathcal{M}^{\mathcal{D}}_{j\beta} \tag{2.3.15}$$

where

$$\Pi^{\mathcal{D}}_{\beta j} = \frac{1}{[16 V E_N E_{N'} E_\beta(k') E_j(q)]^{\frac{1}{2}}} \tag{2.3.16}$$

$$\mathcal{M}^{\mathcal{D}}_{j\beta} = \sqrt{2} G_F\, \overline{\mathcal{V}}_{j,h'}(\vec{q})\, \gamma^\mu \mathbb{L}\, \mathcal{V}_{\beta,h_f}(\vec{k}')\, j^{N,N'}_\mu(P) \quad ; \quad \vec{q} = \vec{k} - \vec{p} = \vec{P} + \vec{k}'\,. \tag{2.3.17}$$

Therefore, the total transition amplitude from production to detection is given by

$$\mathcal{T}_{i\to f} = \sum_{j,h'} U_{\alpha j}\, \Pi^{\mathcal{P}}_{\alpha j}\, \mathcal{M}^{\mathcal{P}}_{\alpha j}\, \mathcal{F}_{\alpha j}\, G_{\beta j}\, U^*_{\beta j}\, \Pi^{\mathcal{D}}_{\beta j}\, \mathcal{M}^{\mathcal{D}}_{j\beta} \tag{2.3.18}$$

[3]This matrix element may be written in terms of vector and axial vector form factors, but such expansion is not necessary in our analysis.

where we have suppressed all the arguments to simplify notation. The factors $\mathcal{F}_{\alpha j}$ and $G_{\beta j}$ encode the *time dependence* of the production, measurement of the charged lepton produced with the (anti) neutrino and final detection processes and the energy uncertainty from the finite lifetime of the parent meson. As noted above (see eqn.2.3.4), $\mathcal{F}_{\alpha j}$ describes nearly energy conservation in the long time narrow width limit but includes the energy uncertainty from the width of the decaying state. Similarly

$$G_{\beta j} \stackrel{t\to\infty}{\to} 2\pi\delta(E_F - E_N - E_j) \tag{2.3.19}$$

describes energy conservation at the detection vertex in the long time limit.

The phases in these factors encode the information of interference effects between the different mass eigenstates.

In order to isolate the contribution of these factors there are several approximations that are dictated by the experimental aspects:

Approximations:

1. For neutrino masses consistent with oscillation experiments utilizing baselines of a few hundred meters, namely $m_j \simeq$ eV, and typical neutrino energy from meson decay, $\gtrsim$ 30 MeV, the neutrinos are ultrarelativistic so we can approximate $E_j(q) = E(q) + m_j^2/2E(q)$ with $E(q) = q$. Obviously, this approximation is valid for even higher energies and longer baselines so that the results may be extrapolated appropriately.
2. We neglect the neutrino masses in the factors $E_j(q)$ in the denominators in $\Pi^{\mathcal{P}}_{\alpha j}\,, \Pi^{\mathcal{D}}_{\beta j}$ (eqns.2.3.2,2.3.16).
3. We also neglect the mass dependence of neutrino spinors $\mathcal{V}$, which depend upon the mass through the factor $\varepsilon_j(q) = m_j/(E_j(q) + q)$ (see (2.8.8,2.8.9)). Neglecting the neutrino masses in the spinors leads to the production and detection matrix elements $\mathcal{M}^{\mathcal{P}}\,, \mathcal{M}^{\mathcal{D}}$ to be independent of the neutrino masses, therefore independent of the label j.
4. Neglecting the neutrino mass, the negative chirality (anti) neutrino only features a positive helicity component, therefore only $h' = +$ remains in the sum. This is, obviously, a consequence of $\varepsilon_j \ll 1$.

Under these approximations and the unitarity of the mixing matrix U, the normalization $\mathcal{N}_\nu(q)$ (2.3.6) of the neutrino state $|\widetilde{\overline{\nu}(\vec{q})}\rangle$ becomes

$$\mathcal{N}_\nu(q) = \frac{\left|\Pi^{\mathcal{P}}_\alpha \, \mathcal{M}^{\mathcal{P}}_\alpha\right|^2}{\Gamma_M(p)} \left[1 - e^{-\Gamma_M(p)t_c}\right] 2\pi\,\delta\big(\mathcal{E}_{\mathcal{P}}\big) \;\; ; \;\; \mathcal{E}_{\mathcal{P}} = E_M - E_\alpha - E\,. \tag{2.3.20}$$

The time dependent factors $\mathcal{F}_{\alpha j}, G_{\beta j}$ feature phases whose interference leads to the oscillations in the transition probabilities, therefore the terms $m_j^2/2E(q)$ must be kept in these phases.

Under these approximations, the factors $\Pi^{\mathcal{P}}\mathcal{M}^{\mathcal{P}}$ and $\Pi^{\mathcal{D}}\mathcal{M}^{\mathcal{D}}$ can be taken out of the sum and the final result for the transition amplitude *factorizes* into production, propagation with oscillations, and detection contributions:

$$\mathcal{T}_{i\to f} = \underbrace{\left[\Pi^{\mathcal{P}}_\alpha \, \mathcal{M}^{\mathcal{P}}_\alpha\right]}_{\text{Production}} \underbrace{\left[\sum_j U_{\alpha j}\, \mathcal{F}_{\alpha j}\, G_{\beta j}\, U^*_{\beta j}\right]}_{\text{Propagation}-\text{Oscillations}} \underbrace{\left[\Pi^{\mathcal{D}}_\beta \, \mathcal{M}^{\mathcal{D}}_\beta\right]}_{\text{Detection}} \tag{2.3.21}$$

The transition probability is given by

$$|\mathcal{T}_{i\to f}|^2 = \left|\Pi^{\mathcal{P}}_\alpha \, \mathcal{M}^{\mathcal{P}}_\alpha\right|^2 \left[\sum_j \sum_i U_{\alpha j} U^*_{\alpha i}\, \mathcal{F}_{\alpha j} \mathcal{F}^*_{\alpha i}\, G_{\beta j} G^*_{\beta i}\, U^*_{\beta j} U_{\beta i}\right] \left|\Pi^{\mathcal{D}}_\beta \, \mathcal{M}^{\mathcal{D}}_\beta\right|^2 \tag{2.3.22}$$

where $\mathcal{F}_{\alpha j} \equiv \mathcal{F}_{\alpha j}[\vec{k}, \vec{q}, t_c]$ is given by (2.3.3) evaluated at $t = t_c$ and $G_{\beta j}$ is given by (2.3.12).

It proves convenient to introduce:

$$\mathcal{E}_{\mathcal{P}} = E_M(p) - E_\alpha(k) - E(q) \;\; ; \;\; \mathcal{E}_{\mathcal{D}} = E_F - E_N - E(q)\,. \tag{2.3.23}$$

For $m_j \ll E(q)$ and narrow width $\Gamma_M \ll E_M$, the products $\mathcal{F}_{\alpha j}\mathcal{F}^*_{\alpha i}$ are sharply peaked at $\mathcal{E}_{\mathcal{P}}$, becoming nearly energy conserving delta functions in the long time and small width limit (see (2.3.4)). Similarly, $G_{\beta j}G^*_{\beta i}$ is sharply peaked at $\mathcal{E}_{\mathcal{D}}$. Each term $\mathcal{F}, G$ describe approximate energy conservation at the production and detection vertices respectively. In order to extract the coefficients of the energy conserving $\delta(\mathcal{E}_{\mathcal{P}}), \delta(\mathcal{E}_{\mathcal{D}})$, we integrate the respective products with a smooth initial and final density of states that are insensitive to Γ_M and Δ_j (for details see ref.[85]). We find

$$\mathcal{F}_{\alpha j}\mathcal{F}^*_{\alpha i} = \frac{2\pi}{\Gamma_M(p)} \frac{\left[1 - e^{-i\Delta_{ij}t_c}\, e^{-\Gamma_M(p)t_c}\right]}{1 + i\mathcal{R}_{ij}}\, \delta(\mathcal{E}_{\mathcal{P}}) \tag{2.3.24}$$

were we have introduced

$$\Delta_{ij} = \frac{\delta m_{ij}^2}{2E(q)} \quad ; \quad \mathcal{R}_{ij} = \frac{\Delta_{ij}}{\Gamma_M(p)} = \frac{\delta m_{ij}^2}{2\Gamma_M M_M} \, \frac{E_M(p)}{E(q)} \quad ; \quad \delta m_{ij}^2 = m_i^2 - m_j^2 \,, \tag{2.3.25}$$

similarly

$$G_{\beta j} G_{\beta i}^* = 2\pi \, i \, e^{i\Delta_{ij} t_c} \, \frac{\left[1 - e^{i\Delta_{ij}(t_D - t_c)}\right]}{\Delta_{ij}} \, \delta(\mathcal{E}_{\mathcal{D}}) \,. \tag{2.3.26}$$

As usual, one is interested in obtaining the transition *rate*; therefore, we focus on $\frac{d}{dt_D}|\mathcal{T}_{i\to f}|^2$ for which we need

$$\frac{d}{dt_D}\Big(G_{\beta j} G_{\beta i}^*\Big) = 2\pi \, \delta(\mathcal{E}_{\mathcal{D}}) \, e^{i\Delta_{ij} t_D} \tag{2.3.27}$$

Separating the diagonal $i = j$ from off-diagonal terms in the sums in (2.3.22), and using the result (2.3.20) for the normalization of the (anti) neutrino state, we find the *transition rate*

$$\frac{d}{dt_D}|\mathcal{T}_{i\to f}|^2 = \Big[\mathcal{N}_\nu\Big] \, \mathcal{P}_{\alpha\to\beta} \left[\frac{d\Gamma_{\overline{\nu}N\to l_\beta N'}}{(2\pi)^6 V^2 \, d^3k' d^3P}\right] \tag{2.3.28}$$

where

$$\frac{d\Gamma_{\overline{\nu}N\to l_\beta N'}}{(2\pi)^6 V^2 \, d^3k' d^3P} = \Big|\Pi_\beta^{\mathcal{D}} \, \mathcal{M}_\beta^{\mathcal{D}}\Big|^2 \, 2\pi \, \delta(\mathcal{E}_{\mathcal{D}}) \tag{2.3.29}$$

is the double differential detection rate for $\overline{\nu}N \to l_\beta^+ N'$ for an incoming *massless* neutrino, and $\mathcal{P}_{\alpha\to\beta}$ is the flavor transition probability

$$\mathcal{P}_{\alpha\to\beta} = \sum_{j,i} U_{\alpha j} U_{\beta j}^* U_{\alpha i}^* U_{\beta i} I_{ij} \tag{2.3.30}$$

where I_{ij} are the interference terms

$$I_{ij} = e^{i\Delta_{ij} t_D} \left[\frac{1 - e^{-i\Delta_{ij} t_c} \, e^{-\Gamma_M(p) t_c}}{1 - e^{-\Gamma_M(p) t_c}}\right] \left[\frac{1 - i\mathcal{R}_{ij}}{1 + \mathcal{R}_{ij}^2}\right] \quad ; \quad I_{ji} = I_{ij}^* \,. \tag{2.3.31}$$

Unitarity of the U matrix allows to write

$$\begin{aligned} \mathcal{P}_{\alpha\to\beta} = \delta_{\alpha,\beta} \; - \quad & 2 \sum_{j>i} \mathrm{Re}\Big[U_{\alpha j} U_{\beta j}^* U_{\alpha i}^* U_{\beta i}\Big] \, \mathrm{Re}\Big[1 - I_{ij}\Big] \\ & -2 \sum_{j>i} \mathrm{Im}\Big[U_{\alpha j} U_{\beta j}^* U_{\alpha i}^* U_{\beta i}\Big] \, \mathrm{Im}\Big[I_{ij}\Big] \,. \end{aligned} \tag{2.3.32}$$

In the above expressions we have implicitly assumed Dirac neutrinos, the case of Majorana neutrinos is obtained by the replacement (see eqns. (2.2.3,2.2.4)) $U \to \widetilde{U}$; $\widetilde{U}_{\alpha j} = U_{\alpha j}\, e^{i\phi_j/2}\, \forall \alpha$ from which it is obvious that the CP-violating Majorana phases do not play any role in $\overline{\nu}_\alpha \to \overline{\nu}_\beta$ oscillations.

The possibility of CP violation in the neutrino sector from Dirac phases is encoded in the imaginary part in (2.3.32) since for the transition probabilities for $\nu_\alpha \to \nu_\beta$ it follows that $U_{\alpha i} \to U^*_{\alpha i}$. Therefore decoherence effects in the *imaginary part* of I_{ij} lead to possible *suppression of CP-violating contributions in the transition probabilities.*

The transition rate (2.3.28), along with (2.3.32), are some of the important results of this article; the factorized form of (2.3.28) is a consequence of the approximations described above. The origin of the prefactor $\mathcal{N}_\nu$ is clear, it is the normalization of the neutrino state that emerges from disentangling the charged lepton in the production process, since this is the correct neutrino state that propagates to the detector and triggers the charged current reaction that yields the measured charged lepton in the final state. The interference terms (2.3.31) encode the decoherence effects arising from the finite lifetime of the source *and* the energy uncertainty associated with the time scale in which the charged lepton produced in a correlated quantum state with the neutrino is observed (or captured). This decoherence can be understood clearly in two limits:

- When $\Delta_{ij} \ll \Gamma_M$ it follows that $\mathcal{R}_{ij} \to 0$ and I_{ij} is the usual interference term. In this limit the energy uncertainty associated with the lifetime of the source does not allow to separate the mass eigenstates and the coherence of the superposition of mass eigenstates is maintained. However, in the opposite limit, $\Delta_{ij} \gg \Gamma_M$, the factor $\mathcal{R}_{ij} \gg 1$ and the interference term is suppressed. In this limit, the lifetime of the source is long, the corresponding energy uncertainty is small and the mass eigenstates are separated in the time evolution and coherence between them in the superposition is suppressed.
- In the limit $\Gamma_M \to 0$ it follows that

$$I_{ij} \to e^{i\Delta_{ij}(t_D - t_c/2)}\, \frac{\sin[\Delta_{ij} t_c/2]}{[\Delta_{ij} t_c/2]}\,. \tag{2.3.33}$$

There are two effects in this expression: 1) a shortening of the baseline by the distance

travelled by the charged lepton produced with the (anti) neutrino and 2) a suppression factor associated with the time uncertainty: if $\Delta_{ij} t_c > 1$, then the interference term is suppressed, this is because if the charged lepton produced with the (anti) neutrino is entangled all throughout the evolution at long time $t_c \gg 1/\Delta_{ij}$ the energy uncertainty becomes much smaller than the difference in energy between mass eigenstates and these are projected out by energy conservation which leads to their decoherence in the superposition. This is another manifestation of energy conservation as encoded in Fermi's Golden rule. In terms of the oscillation length, L_{ij}^{osc}, defined as

$$\Delta_{ij} = \frac{\delta m_{ij}^2}{2E} \equiv \frac{2\pi}{L_{ij}^{osc}} \tag{2.3.34}$$

the suppression factor $2\sin[\Delta_{ij} t_c/2]/\Delta_{ij} t_c < 1$ when the stopping length scale $L_c \equiv t_c \simeq L_{ij}$. The case $\Gamma \to 0$ is relevant for reactor experiments. See the discussion in section(2.6.1).

The suppression factor associated with the lifetime is relevant in the case of possible new generation of (sterile) neutrinos with masses in the eV range when produced in the decay of pions or kaons.

For pion decay at rest, the typical energy of a (nearly massless) neutrino is $E^* \sim 30\,\text{MeV}$, the pion width at rest $\Gamma_\pi = 2.5 \times 10^{-8}\,\text{eV}$ and for one generation of sterile neutrino with $m_4 \gg m_{1,2,3}$ we find

$$\mathcal{R} \simeq \frac{m_4^2}{2E^*\,\Gamma_\pi} \simeq \frac{2}{3}\Big(\frac{m_4}{\text{eV}}\Big)^2, \tag{2.3.35}$$

therefore, for $m_4 \geq 1\,\text{eV}$, the suppression factor can be substantial and the transition probability is suppressed. For the decay of a pion in flight with a large Lorentz γ factor, the result only changes by a factor 2 as can be seen as follows: consider a neutrino that is emitted collinear with the direction of the pion in the laboratory frame (say along the $z - axis$), its energy in the laboratory frame is

$$E = \gamma E^* \big(1 + V_\pi\big) \tag{2.3.36}$$

where V_π is the pion's velocity, for $\gamma \gg 1$ it follows that $E \sim 2\gamma E^*$. The width of the pion in the laboratory frame is Γ_π/γ; therefore, for neutrinos produced by pion decay in flight with a large Lorentz factor

$$\mathcal{R} \simeq \frac{1}{3}\left(\frac{m_4}{\mathrm{eV}}\right)^2 . \tag{2.3.37}$$

In conclusion, for new generations of (sterile) neutrinos with masses in the eV range, experiments in which oscillations are probed with neutrinos from pion decay feature the suppression factors associated with the pion width. For Kaons, the situation improves because in this case

$$\Gamma_K \simeq 5 \times 10^{-8}\,\mathrm{eV} \quad ; \quad E^* \simeq 235.5\,\mathrm{MeV}$$

and for Kaon (DAR)

$$\mathcal{R} \simeq \frac{1}{25}\left(\frac{m_4}{\mathrm{eV}}\right)^2 ,$$

thus $\mathcal{R} < 1$ for $m_4 \simeq$ few eV.

2.4 $3+2$ AND $3+1$ CASES IN THE "SHORT-BASELINE APPROXIMATION":

In the "short-baseline" approximation, we assume that there are sterile neutrinos $j = 4, 5 \cdots$ with $m_4, m_5 \cdots \gg m_1, m_2, m_3$ so that $\delta m^2 L/E \simeq \mathcal{O}(1)$ for $L \simeq 10 - 1000\,\mathrm{mts}$ corresponding to short baseline experiments.

We begin by considering the $3+2$ scenario from which we will extract the case $3+1$.

3+2 case: In this case, $m_5, m_4 \gg m_1, m_2, m_3$ so that

$$I_{ij} \simeq 1 \quad , \; i,j = 1,2,3 \quad ; \quad I_{i4} = I_{14} \quad ; \quad I_{i5} = I_{15} \quad , \quad i = 1,2,3\,. \tag{2.4.1}$$

Unitarity of the U matrix entails

$$\sum_{i=1}^{3} U^*_{\alpha i} U_{\beta i} = \delta_{\alpha\beta} - U^*_{\alpha 4} U_{\beta 4} - U^*_{\alpha 5} U_{\beta 5}\,. \tag{2.4.2}$$

Separating the terms with $j = 4, 5$ in (2.3.32), we find for $\alpha \neq \beta$ (appearance)

$$\begin{aligned}
\mathcal{P}_{\alpha\to\beta} &= 4|U_{\alpha4}||U_{\beta4}|\Big[|U_{\alpha4}||U_{\beta4}| + |U_{\alpha5}||U_{\beta5}|\cos\phi_{54}\Big]\frac{1}{2}\,\mathrm{Re}[1 - I_{41}] \\
&- 4|U_{\alpha4}||U_{\beta4}||U_{\alpha5}||U_{\beta5}|\cos\phi_{54}\,\frac{1}{2}\,\mathrm{Re}[1 - I_{54}] \\
&+ 4|U_{\alpha5}||U_{\beta5}|\Big[|U_{\alpha5}||U_{\beta5}| + |U_{\alpha4}||U_{\beta4}|\cos\phi_{54}\Big]\frac{1}{2}\,\mathrm{Re}[1 - I_{51}] \\
&+ 2\Big[|U_{\alpha4}||U_{\beta4}||U_{\alpha5}||U_{\beta5}|\sin\phi_{54}\Big]\mathrm{Im}\Big[\mathrm{I}_{41} - \mathrm{I}_{51} + \mathrm{I}_{54}\Big]\,. \qquad (2.4.3)
\end{aligned}$$

where following [59] we have defined

$$\phi_{54} = \mathrm{Arg}\Big[U_{\alpha5}U^*_{\beta5}U^*_{\alpha4}U_{\beta4}\Big]\ ,\ \text{for}\ \alpha = e\,,\beta = \mu\,, \qquad (2.4.4)$$

and used $\mathrm{Im}[I_{ij}] = -\mathrm{Im}[I_{ji}]$. We note that interchanging $\alpha \leftrightarrow \beta$ ($e \leftrightarrow \mu$) is equivalent to the exchange $4 \leftrightarrow 5$, namely $\phi_{54} \to -\phi_{54} = \phi_{45}$ which leaves the result (2.4.3) invariant since $\mathrm{Re}[I_{ji}]$ is even and $\mathrm{Im}I_{ji}$ odd respectively under $i \leftrightarrow j$. If $\phi_{54} \neq 0$, there is CP-violation in the neutrino sector because $\phi_{54} \to -\phi_{54}$ for $\nu \to \nu$ oscillations since this implies that the elements of the mixing matrix $U_{\alpha i} \to U^*_{\alpha i}$.

The 3+2 case effectively describes mixing between *three* species; consequently, it features only one *effective* CP-violating angle.

For $\alpha = \beta$ (disappearance), we find

$$\begin{aligned}
\mathcal{P}_{\alpha\to\alpha} &= 1 - 4\Big\{|U_{\alpha4}|^2\Big[1 - |U_{\alpha4}|^2 - |U_{\alpha5}|^2\Big]\frac{1}{2}\,\mathrm{Re}[1 - I_{41}] \\
&+ |U_{\alpha4}|^2|U_{\alpha5}|^2\,\frac{1}{2}\,\mathrm{Re}[1 - I_{54}] \\
&+ |U_{\alpha5}|^2\Big[1 - |U_{\alpha4}|^2 - |U_{\alpha5}|^2\Big]\frac{1}{2}\,\mathrm{Re}[1 - I_{51}]\Big\}\,, \qquad (2.4.5)
\end{aligned}$$

which does not feature a contribution from the CP-violating angle.

3+1 case: This case is obtained from the 3+2 case above by setting $U_{\alpha5} = 0\,\forall\alpha$, leading to the appearance probability ($\alpha \neq \beta$) ,

$$\mathcal{P}_{\alpha\to\beta} = 4|U_{\alpha4}|^2|U_{\beta4}|^2\,\frac{1}{2}\,\mathrm{Re}[1 - I_{41}]\,, \qquad (2.4.6)$$

and the disappearance (survival) probability ($\alpha = \beta$)

$$\mathcal{P}_{\alpha\to\alpha} = 1 - 4|U_{\alpha 4}|^2\Big[1 - |U_{\alpha 4}|^2\Big]\,\frac{1}{2}\,\mathrm{Re}[1 - I_{41}]\,. \tag{2.4.7}$$

The $3+1$ case effectively describes mixing between two generations and, consequently, does not feature any CP-violating contribution. For this case, it is often convenient[38] to introduce the effective mixing angles

$$\sin^2 2\theta_{\alpha\beta} \equiv 4|U_{\alpha 4}|^2|U_{\beta 4}|^2 \quad , \quad \alpha \neq \beta \tag{2.4.8}$$

$$\sin^2 2\theta_{\alpha\alpha} \equiv 4|U_{\alpha 4}|^2\Big[1 - |U_{\alpha 4}|^2\Big] \quad , \quad \alpha = \beta\,. \tag{2.4.9}$$

2.5 MAJORANA STERILE NEUTRINOS AND $|\Delta L| = 2$ $\nu \leftrightarrow \overline{\nu}$ OSCILLATIONS:

In the previous section we have assumed that sterile neutrinos are of the Dirac variety; however, if neutrinos are Majorana fermions, new processes, such as neutrino-less double beta decay (see[107] for recent reviews) and $\nu \leftrightarrow \overline{\nu}$ oscillations, are available. As discussed in ref.[88], $\nu \leftrightarrow \overline{\nu}$ oscillations have the potential to reveal CP-violating Majorana phases and, to make clear the Majorana nature of the mixing matrix, we write it as $\widetilde{U}$ following eqn. (2.2.3).

These processes can be understood by considering the full interaction Hamiltonian including the the hermitian conjugate of the one displayed in (2.3.13), namely

$$\begin{aligned} H_i &= \widetilde{U}^*_{\beta j}\,\sqrt{2}\,G_F \int \overline{\Psi}_{\nu_j}(\vec{x})\gamma^\mu\mathbb{L}\,\Psi_{l_\beta}(\vec{x})\,\mathcal{J}^{(N,N')}_\mu(\vec{x})d^3x \\ &+ \widetilde{U}_{\beta j}\,\sqrt{2}\,G_F \int \overline{\Psi}_{l_\beta}(\vec{x})\gamma^\mu\mathbb{L}\,\Psi_{\nu_j}(\vec{x})\,\mathcal{J}^{\dagger\,(N,N')}_\mu(\vec{x})d^3x\,. \end{aligned} \tag{2.5.1}$$

The first line yields the $\Delta L = 0$ $\overline{\nu} \leftrightarrow \overline{\nu}$ oscillations just as for the Dirac case discussed in the previous section. The second line contributes to the detection process *only* for Majorana neutrinos and yields the $|\Delta L| = 2$ contribution, as can be simply understood from the

following argument pertaining to Majorana fermions: the production Hamiltonian (2.2.1) is determined by charge conservation: a π^- decays into a negatively charged lepton l^-_α, thus requiring the $\overline{\Psi}_{l_\alpha}$ in (2.2.1), the Ψ_{ν_j} *creates* a neutrino (same as an *antineutrino* for Majoranas) with an operator $\hat{b}^\dagger_{\vec{k},h}$ that multiplies a charge conjugate spinor $\mathcal{U}^c_h(\vec{k})$ (see the expansion (2.8.17) in appendix 2.8.1). Using the first line in (2.5.1), the neutrino is destroyed at the detection vertex using a $\hat{b}_{\vec{k},h}$ of the $\overline{\Psi}_{\nu_j}$ which *also* multiplies the spinor $\mathcal{U}^c_h(\vec{k})$ along with the creation of positively charged lepton l^+_β. Therefore, this $\Delta L = 0$ contribution is the same as that for a Dirac neutrino and features the product $\widetilde{U}_{\alpha j}\widetilde{U}^*_{\beta j}$, which is insensitive to the Majorana phase. However, the neutrino in the intermediate state can also be annihilated by using a $\hat{b}_{\vec{k},h}$ from Ψ_{ν_j} in the second line, which now multiplies the spinor $\mathcal{U}_h(\vec{k})$, along with the creation of a *negatively charged lepton* l^-_β from $\overline{\Psi}_\beta$. This contribution features the product $\widetilde{U}_{\alpha j}\widetilde{U}_{\beta j} = U_{\alpha j}U_{\beta j}\, e^{i\theta_j}$ and manifestly displays the Majorana phase. This process is depicted in fig.(3).

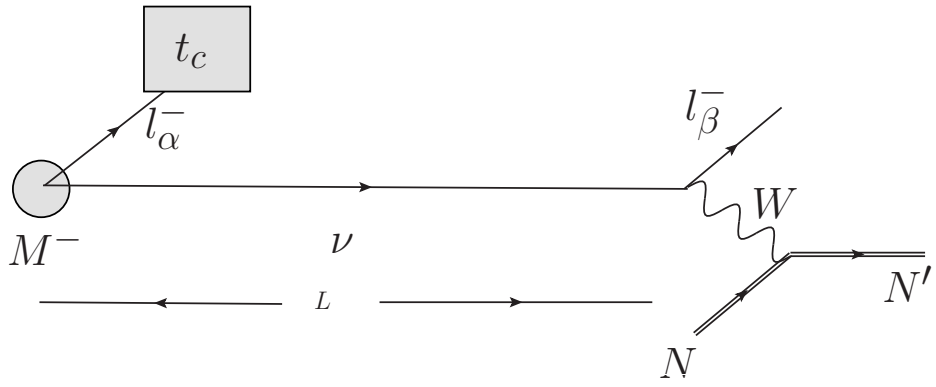


Figure 3: $|\Delta L| = 2$ process from Majorana neutrinos. The charged lepton l^-_α produced with the neutrino is observed, absorbed or decays at a time t_c and another charged lepton l^-_β is detected.

The CP-conjugate process $\pi^+ \to l^+_\alpha\, \nu \to \nu N \to N'\, l^+_\beta$ with $|\Delta L| = 2$ features the product $\widetilde{U}^*_{\alpha j}\widetilde{U}^*_{\beta j}$ showing that the Majorana phase is also CP-violating. It is convenient to introduce

$$\widetilde{\sigma}^\mu = (1, -\vec{\sigma}) \tag{2.5.2}$$

we now find for the transition matrix element $\mathcal{T}_{i\to f}$ from the initial ($i = \pi^- N$) to the final ($f = l_\alpha^- l_\beta^- N'$) state

$$\begin{aligned}\mathcal{T}_{i\to f} &= \sqrt{2} F_M G_F \sum_h \sum_j \widetilde{U}_{\alpha j} \widetilde{U}_{\beta j} \left(\mathcal{U}^\dagger_{l_\alpha, h'}\right)_L \left(\widetilde{\sigma}\cdot J^M\right) \left(\mathcal{U}^c_{h,j}(\vec{q})\right)_L \mathcal{F}_{\alpha j}\, \Pi^{\mathcal{P}}_{\alpha j} \\ &\times \left(\mathcal{U}^\dagger_{l_\beta, h''}\right)_L \left(\widetilde{\sigma}\cdot J^{(N,N')}\right) \left(\mathcal{U}_{h,j}(\vec{q})\right)_L G_{\beta j}\, \Pi^{\mathcal{D}}_{\beta j} \end{aligned} \tag{2.5.3}$$

where again we suppressed arguments to simplify notation. The sum over helicity states h can be carried out straightforwardly using the results of appendix (2.8.1), we find (no sum over α, β)

$$\mathcal{T}_{i\to f} = \left[T^{\alpha\beta}_{-+} - T^{\alpha\beta}_{+-}\right] \sum_j \widetilde{U}_{\alpha j} \widetilde{U}_{\beta j} \frac{m_j}{2E_j(q)} \mathcal{F}_{\alpha j}\, G_{\beta j}\,, \tag{2.5.4}$$

where we have introduced

$$\begin{aligned} T^{\alpha\beta}_{ab} &\equiv \sqrt{2} F_M G_F\, 2E(q)\, \Pi^{\mathcal{P}}_\alpha\, \Pi^{\mathcal{D}}_\beta \left[\left(\mathcal{U}^\dagger_{l_\alpha, h'}\right)_L \left(\widetilde{\sigma}\cdot J^M\right) v_a(\vec{q})\right] \left[\left(\mathcal{U}^\dagger_{l_\beta, h''}\right)_L \left(\widetilde{\sigma}\cdot J^{(N,N')}\right) v_b(\vec{q})\right] \\ & a, b = -, + \end{aligned} \tag{2.5.5}$$

where the Weyl spinors $v_\pm(q)$ are the helicity eigenstates (2.8.12). Here, $E(q) = q$ for *massless neutrinos* and $T^{\alpha\beta}_{ab}$ do not depend on the mass eigenstate label j. In arriving at expressions (2.5.4,2.5.5) we have written $\Pi^{\mathcal{P}}_{\alpha j}\, \Pi^{\mathcal{D}}_{\beta j} = \Pi^{\mathcal{P}}_\alpha\, \Pi^{\mathcal{D}}_\beta\, 2E(q)/2E_j(q)$ where the $\Pi^{\mathcal{P}}_\alpha\, \Pi^{\mathcal{D}}_\beta$ now correspond to the phase space factors (2.3.2,2.3.16) for *massless neutrinos*, namely $E_j(q) \to E(q) = q$.

The amplitudes $T^{\alpha\beta}_{ab}$ have a simple interpretation: $T^{\alpha\beta}_{-+}$ is the amplitude for the combined process $\pi^- \to l^-_\alpha \nu$, $\overline{\nu} N \to l^-_\beta N'$ and $T^{\alpha\beta}_{+-}$ for the process $\pi^- \to l^-_\alpha \overline{\nu}$, $\nu N \to l^-_\beta N'$ where $\nu, \overline{\nu}$ are *massless* left handed neutrinos (and right handed antineutrinos), corresponding to the $\nu \leftrightarrow \overline{\nu}$ mixing that violates lepton number by two units. These two amplitudes contribute coherently to the process $\pi^- N \to l^-_\alpha l^-_\beta N'$ and are added (with their respective signs) in the total amplitude. The mass dependence in the transition amplitude is a consequence of a helicity change in the $\Delta L = 2$ process $\nu \leftrightarrow \bar{\nu}$.

The expression (2.5.4) is generally valid for arbitrary masses of Majorana sterile neutrinos and, with a simple modification of the final state, also describes the $\Delta L = 2$ processes studied in ref.[222].

Proceeding as in the $\Delta L = 0$ case, we finally find for the transition *rate*

$$\frac{d}{dt_D}|\mathcal{T}_{i\to f}|^2 = \Upsilon^{\alpha\beta}\, P^{|\Delta L|=2}_{\alpha\to\beta} \quad , \quad \text{no sum over } \alpha,\beta \tag{2.5.6}$$

where

$$P^{|\Delta L|=2}_{\alpha\to\beta} = \sum_{j,i} \widetilde{U}_{\alpha j}\widetilde{U}_{\beta j}\widetilde{U}^*_{\alpha i}\widetilde{U}^*_{\beta i}\,\frac{m_j m_i}{4E_jE_i}\,I_{ij}\,, \tag{2.5.7}$$

is the $\nu \leftrightarrow \overline{\nu}$ transition probability with $|\Delta L| = 2$ and

$$\Upsilon^{\alpha\beta} = \left|T^{\alpha\beta}_{-+} - T^{\alpha\beta}_{+-}\right|^2 \frac{\left[\Pi^{\mathcal{P}}_\alpha\right]^2}{\Gamma_M(p)}\left[1 - e^{-\Gamma_M(p)t_c}\right]\, 2\pi\,\delta\big(\mathcal{E}_{\mathcal{P}}\big)\left[\Pi^{\mathcal{D}}_\beta\right]^2 2\pi\,\delta(\mathcal{E}_{\mathcal{D}}) \tag{2.5.8}$$

encodes the transition matrix elements for production and detection. We note that unlike the $\Delta L = 0$ case, here there is *no factorization* of production and detection, this is a consequence of the fact that $\nu \leftrightarrow \overline{\nu}$ oscillation implies helicity change (and a mass insertion) and both helicity changing contributions contribute coherently to the total amplitude as explained above. A similar observation was pointed out in ref.[108]. We are not concerned here with $\Upsilon^{\alpha\beta}$ but with the transition probability $P^{|\Delta L|=2}_{\alpha\beta}$, which can be written as

$$\begin{aligned} P^{|\Delta L|=2}_{\alpha\to\beta} = \sum_j |U_{\alpha j}|^2\,|U_{\beta j}|^2\,\frac{m_j^2}{4E_j^2} \;&+\; 2\sum_{j>i}\mathrm{Re}[\widetilde{U}_{\alpha j}\widetilde{U}_{\beta j}\widetilde{U}^*_{\alpha i}\widetilde{U}^*_{\beta i}]\,\frac{m_j m_i}{4E_jE_i}\,\mathrm{Re}[I_{ji}] \\ &+\; 2\sum_{j>i}\mathrm{Im}[\widetilde{U}_{\alpha j}\widetilde{U}_{\beta j}\widetilde{U}^*_{\alpha i}\widetilde{U}^*_{\beta i}]\,\frac{m_j m_i}{4E_jE_i}\,\mathrm{Im}[I_{ji}]\,, \end{aligned} \tag{2.5.9}$$

Just as in the $\Delta L = 0$ case, the main difference with the usual quantum mechanical case is the replacement

$$e^{i\Delta_{ij}L} \to I_{ij} = e^{i\Delta_{ij}L}\left[\frac{1 - e^{-i\Delta_{ij}L_c}\,e^{-\Gamma_M(p)L_c}}{1 - e^{-\Gamma_M(p)L_c}}\right]\left[\frac{1 - i\mathcal{R}_{ij}}{1 + \mathcal{R}^2_{ij}}\right], \tag{2.5.10}$$

where $\Delta_{ij}; \mathcal{R}_{ij}$ are given by eqn. (2.3.25) where the extra factors describe decoherence effects associated with the lifetime of the decaying meson and the measurement of the charged lepton partner of the produced neutrino.

To be sure *if* the absolute mass scale of the new generation of sterile neutrinos is $\simeq$ eV then the factor $\simeq m^2/E^2 \lesssim 10^{-14}$ makes the $|\Delta L| = 2$ contribution all but unobservable

with the current (and foreseeable) facilities for short-baseline experiments with $m \simeq$ eV. However, oscillation experiments measure the squared mass *differences*; therefore, in absence of a determination of the absolute scale of masses, there remains the possibility that new generation of sterile neutrinos may be heavy but nearly degenerate so that the difference in squared masses is small and lead to interference and oscillations on the length scales of short baseline experiments and $P_{\alpha\to\beta}^{|\Delta L|=2}$ is not negligible .

3+2 and 3+1 schemes: Under the assumption that $m_4, m_5 \gg m_i, i = 1, 2, 3$, the contribution from active-like mass eigenstates is clearly subleading for the $|\Delta L| = 2$ transitions; therefore, keeping only the two largest mass eigenstates

$$\begin{aligned} P_{\alpha\to\beta}^{|\Delta L|=2} &= |U_{\alpha 5}|^2 \, |U_{\beta 5}|^2 \, \frac{m_5^2}{4E_5^2} + |U_{\alpha 4}|^2 \, |U_{\beta 4}|^2 \, \frac{m_4^2}{4E_4^2} + \cdots \\ &+ 2|U_{\alpha 5}||U_{\beta 5}||U_{\alpha 4}||U_{\beta 4}| \, \cos(\delta_{54} + \theta_{54}) \, \frac{m_5 m_4}{4E_5 E_4} \, \mathrm{Re}[I_{54}] + \cdots \\ &+ 2|U_{\alpha 5}||U_{\beta 5}||U_{\alpha 4}||U_{\beta 4}| \, \sin(\delta_{54} + \theta_{54}) \, \frac{m_5 m_4}{4E_5 E_4} \, \mathrm{Im}[I_{54}] + \cdots \end{aligned} \tag{2.5.11}$$

where the dots stand for the contributions from $i = 1, 2, 3$, U is the Dirac mixing matrix (2.2.3) and

$$\delta_{54} = \mathrm{Arg}\Big[U_{\alpha 5} U_{\beta 5} U_{\alpha 4}^* U_{\beta 4}^*\Big] \ , \ \text{for } \alpha = e \, , \beta = \mu \tag{2.5.12}$$

is a *Dirac CP-violating phase different* from the ϕ_{54} that enter in the $\Delta L = 0$ case (2.4.4) and $\theta_{54} = \theta_5 - \theta_4$ with θ_j the Majorana CP-violating phases (2.2.3).

The $3 + 1$ scheme is obtained simply by setting $U_{\alpha 5} = 0 \, \forall \alpha$ in which case there are no oscillations to leading order in m/E.

2.6 ANALYSIS OF DECOHERENCE EFFECTS IN ACCELERATOR EXPERIMENTS:

The decoherence effects associated with the lifetime of the source and the measurement (or capture) length scale of the charged lepton emitted with the (anti) neutrino are encoded

in the quantities $\mathrm{Re}[I_{ji}]$, $\mathrm{Im}[I_{ji}]$ given by eqns. (2.6.3,2.6.4) the latter one determines the suppression of the CP violating contributions from these decoherence effects. In this section we compare these terms to the familiar ones obtained from the quantum mechanical description of neutrino oscillations (2.6.6) as a function of the neutrino energy for fixed baselines.

Introducing

$$\Delta_{ji}(E) = \frac{\delta m^2_{ji}}{2E} \;\; ; \;\; \mathcal{R}_{ji} = \frac{\delta m^2_{ji}}{2E\Gamma_M(p)} \;\; ; \;\; \delta m^2_{ji} = m^2_j - m^2_i \,, \tag{2.6.1}$$

and replacing as usual

$$t_D \to L \;\; ; \;\; t_c \to L_c \tag{2.6.2}$$

we find

$$\begin{aligned} \mathrm{Re}[I_{ji}] \;\; = \;\; & \frac{1}{1+\mathcal{R}^2_{ji}} \, \frac{1}{1-e^{-\Gamma_M(p)L_c}} \left[\left(\cos\left[\frac{\delta m^2_{ji}}{2E}\, L\right] + \mathcal{R}_{ji} \sin\left[\frac{\delta m^2_{ji}}{2E}\, L\right] \right) - \right. \\ & \left. e^{-\Gamma_M(p)L_c} \left(\cos\left[\frac{\delta m^2_{ji}}{2E}\, (L-L_c)\right] + \mathcal{R}_{ji} \sin\left[\frac{\delta m^2_{ji}}{2E}\, (L-L_c)\right] \right) \right] , \end{aligned} \tag{2.6.3}$$

$$\begin{aligned} \mathrm{Im}[I_{ji}] \;\; = \;\; & \frac{1}{1+\mathcal{R}^2_{ji}} \, \frac{1}{1-e^{-\Gamma_M(p)L_c}} \left[\left(\sin\left[\frac{\delta m^2_{ji}}{2E}\, L\right] - \mathcal{R}_{ji} \cos\left[\frac{\delta m^2_{ji}}{2E}\, L\right] \right) - \right. \\ & \left. e^{-\Gamma_M(p)L_c} \left(\sin\left[\frac{\delta m^2_{ji}}{2E}\, (L-L_c)\right] - \mathcal{R}_{ji} \cos\left[\frac{\delta m^2_{ji}}{2E}\, (L-L_c)\right] \right) \right] . \end{aligned} \tag{2.6.4}$$

we note that

$$\frac{\delta m^2_{ji}}{2E}\, L_c \equiv \mathcal{R}_{ji}\, \Gamma_M(p) L_c \tag{2.6.5}$$

this relation highlights that there are only two combination of parameters that determine the corrections, namely $\mathcal{R}_{ji}$ and $\Gamma_M(p)L_c$; furthermore, $\Gamma_M(p)L_c \equiv L_c/l_M(p)$ where $l_M(p)$ is the decay length of the meson in the laboratory frame. We would like to point out that similar results have been obtained in refs[86] in which wave packets are analyzed throughout the production/detection process whereas our results are obtained in a completely different manner. In our treatment, we did not attempt to include localization wavepackets for the pion and, in the WW treatment, the pions would be the only source where an introduction

of wavepackets would be appropriate. The usual decay matrix elements for pion decay from quantum field theory were used and a full discussion of wavepackets is available in refs[86].

In absence of the decoherence contributions, the usual expressions emerge, namely

$$\mathrm{Re}[I_{ji}] = \cos\left[\frac{\delta m^2_{ji}\, L}{2E}\right] \;\; ; \;\; \mathrm{Im}[I_{ji}] = \sin\left[\frac{\delta m^2_{ji}\, L}{2E}\right], \tag{2.6.6}$$

with

$$\frac{\delta m^2_{ji}\, L}{2E} = 2.54\left(\frac{\delta m^2_{ji}}{\mathrm{eV}^2}\right)\left(\frac{L}{\mathrm{km}}\right)\left(\frac{\mathrm{GeV}}{E}\right). \tag{2.6.7}$$

Whereas the length scale L_c is determined by the particular experimental setting and is therefore a parameter, the width of the parent particle is a function of the neutrino energy through the Lorentz factor as follows.

In the rest frame of the decaying meson, its width is Γ_M and the antineutrino (neutrino) is emitted isotropically with an energy $E^*_j = \sqrt{q^{*2} + m^2_j}$ with q^* given by (2.2.13); in the laboratory frame, where the meson is moving with velocity V_M, the width is Γ_M/γ and the energy of an anti (neutrino) collinear with the meson is blue shifted to

$$E = \gamma E^*(1 + V_M) \tag{2.6.8}$$

where we have neglected the mass of the neutrino. Therefore

$$\gamma(E) = \frac{E^2 + E^{*2}}{2EE^*} \;\; ; \;\; E^* < E \tag{2.6.9}$$

hence

$$\mathcal{R}_{ji}(E) = \frac{\delta m^2_{ji}}{4E^*\,\Gamma_M}\left(1 + \frac{E^{*2}}{E^2}\right). \tag{2.6.10}$$

In the analysis below, we focus on neutrinos from Pion decay and the analysis for Kaon decay is similar. Using the Pion decay width, $\Gamma_\pi = 2.5 \times 10^{-8}\,\mathrm{eV}$, as a benchmark, we obtain

$$\mathcal{R}_{ji}(E) = \frac{1}{3}\left(\frac{\delta m^2_{ji}}{\mathrm{eV}^2}\right)\left(\frac{30\,\mathrm{MeV}}{E^*}\right)\left(\frac{\Gamma_\pi}{\Gamma_M}\right)\left(1 + \frac{E^{*2}}{E^2}\right). \tag{2.6.11}$$

An illuminating interpretation of the results (2.6.3,2.6.4) emerges by defining[4]

$$\cos[\theta_{ji}(E)] = \frac{1}{\sqrt{1+\mathcal{R}_{ji}^2}} \quad ; \quad \sin[\theta_{ji}(E)] = \frac{\mathcal{R}_{ji}}{\sqrt{1+\mathcal{R}_{ji}^2}}\,, \tag{2.6.12}$$

in terms of which we find

$$\begin{aligned} \mathrm{Re}[I_{ji}] \;=\; & \frac{1}{\sqrt{1+\mathcal{R}_{ji}^2}}\,\frac{1}{1-e^{-\Gamma_M(p)L_c}}\;* \\ & \left\{\;\cos\left[\frac{\delta m_{ji}^2}{2E}\,L-\theta_{ji}(E)\right]-e^{-\Gamma_M(p)L_c}\,\cos\left[\frac{\delta m_{ji}^2}{2E}\,(L-L_c)-\theta_{ji}(E)\right]\right\} \end{aligned} \tag{2.6.13}$$

$$\begin{aligned} \mathrm{Im}[I_{ji}] \;=\; & \frac{1}{\sqrt{1+\mathcal{R}_{ji}^2}}\,\frac{1}{1-e^{-\Gamma_M(p)L_c}}\;* \\ & \left\{\;\sin\left[\frac{\delta m_{ji}^2}{2E}\,L-\theta_{ji}(E)\right]-e^{-\Gamma_M(p)L_c}\,\sin\left[\frac{\delta m_{ji}^2}{2E}\,(L-L_c)-\theta_{ji}(E)\right]\right\}. \end{aligned} \tag{2.6.14}$$

While the general case must be studied numerically, the limit $\Gamma_M L_c \gg 1$ provides a most clear assessment: as compared to the usual quantum mechanical expression (2.6.6), the decoherence factors result in i) a suppression of the transition probabilities $\simeq 1/\sqrt{1+\mathcal{R}_{ji}^2}$ and ii) an overall energy dependent phase shift $\theta_{ji}(E)$.

For example, for a sterile neutrino mass $m_s \gtrsim 1\,\mathrm{eV} \gg m_{1,2,3}$ from π decay, it follows that $1 \lesssim \mathcal{R}$ thus from (2.6.11,2.6.12) $\pi/4 \lesssim \theta_{ji} \lesssim \pi/2$. Trying to fit the mass (and mixing angles) by using the usual expression (2.6.6) would imply an effective $\delta m^2_{eff} = \delta m^2 - 2E\,\theta(E)/L$. For example, for accelerators experiments with $E \simeq \mathrm{GeV}$, $L \simeq 1\,\mathrm{Km}$, such a fit would lead to $2E\,\theta(E)/L \simeq \mathrm{eV}^2$ and a *large underestimate of the sterile neutrino mass and the mixing and CP-violating angle.*

A similar interpretation holds for the imaginary part (2.6.15), which is associated with CP-violating amplitudes, the suppression factor would result in an *underestimate* of CP-violation if the usual quantum mechanical expression (2.6.6) is used in fitting experimental data. For both cases, if the product $\Gamma_M L_c \gtrsim 1$ then the usual quantum mechanical formulae will not be valid and decoherence effects must be considered.

[4]Note that $\mathrm{sign}(\theta_{ij}) = \mathrm{sign}(\delta m^2_{ij})$.

This simpler case illustrates that for short baseline accelerator experiments in which neutrinos are produced from the decay of pions and are designed to reveal oscillations of new generations of sterile neutrinos with masses in the eV range, the decoherence aspects associated with the pion lifetime and the stopping length scale of the muon comparable to the decay length of the pion may lead to substantial corrections to the quantum mechanical oscillation probabilities. A more reliable assessment is obtained numerically below for different experimental situations and, in these investigations, we focus on sterile mass ranges that are relevant for current accelerator searches rather than masses relevant to structure formation.

MiniBooNE/SciBooNE: For MiniBooNE/SciBooNE, antineutrinos are produced primarily from $\pi^- \to \mu^- \overline{\nu}_\mu$, Pions decay in a decay tunnel $\simeq 50\,\text{mts}$ long and muons are stopped in the "dirt" at a typical distance $\simeq 4\,\text{mts}$ beyond the decay tunnel[5], therefore in this situation $\Gamma_\pi(p) L_c \simeq 1$. The SciBooNE detector is at a distance $L = 100\,\text{mts}$ from the production region, in between the end of the decay tunnel and MiniBooNE, whose detector is at a baseline $L = 540\,\text{mts}$, and the neutrino energy range (for both) is $0.3 \leq E \lesssim 1.6\,\text{GeV}$. Figs. (4-6) show the comparison between the CP-even and odd parts with (modified) and without (QM) the decoherence corrections for MiniBooNE for $m_s = 1, 2, 3\,\text{eV}$ respectively and figs. (7-9) show the same comparison for SciBooNE parameters with $L = 100m$ and same energy range and values of m_s.

These figures confirm the interpretation of the decoherence modifications in terms of an overall suppression of the amplitude and a phase-shift that leads to an offset in the position of the peaks with respect to the quantum mechanical result. Since the mixing angle is extracted from the maximum amplitude of the probability and the mass from the position of the peaks, a fit with the quantum mechanical formula would *underestimate* both the mixing angle and the mass as analyzed above. A similar conclusion applies to the CP-violating angle. For MiniBooNE, the suppression and off-set are small when $\delta m^2 \lesssim 1\,\text{eV}^2$, resulting in an underestimate of about $3 - 5\%$ in amplitude and mass as shown in fig. (4) but is larger at SciBooNE as shown in fig. (7), but for $\delta m^2 = m_s^2 \sim 3\,\text{eV}^2$, fig. (6) for MiniBooNE reveals $\sim 15\%$ suppression in the amplitude with a similar underestimate in the mass (off-set).

[5]D.B. is indebted to William C. Louis III for correspondence clarifying these aspects.

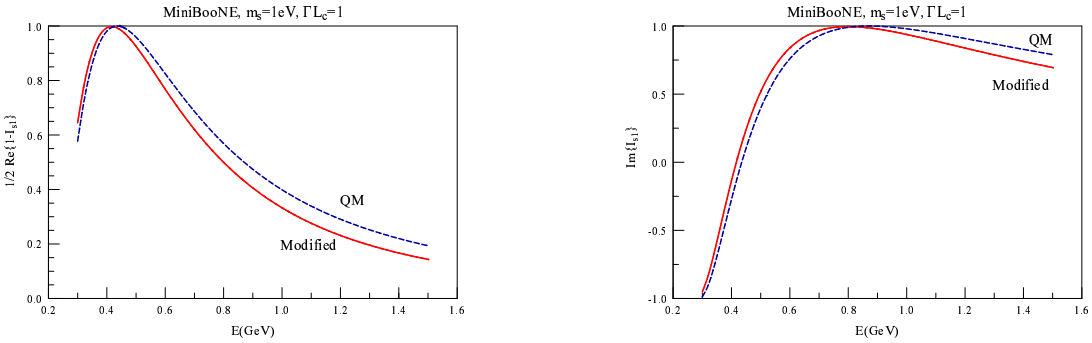


Figure 4: CP-even/odd parts of transition probability for MiniBooNE parameters: $L = 540\text{m}, \Gamma_\pi L_c \simeq 1$ for $m_s = 1\text{eV}$. Solid line (modified) $\text{Re}[1 - I_{s1}]/2$ dashed line (Qm) is the quantum mechanical result $\sin^2[m_s^2/4E]$.

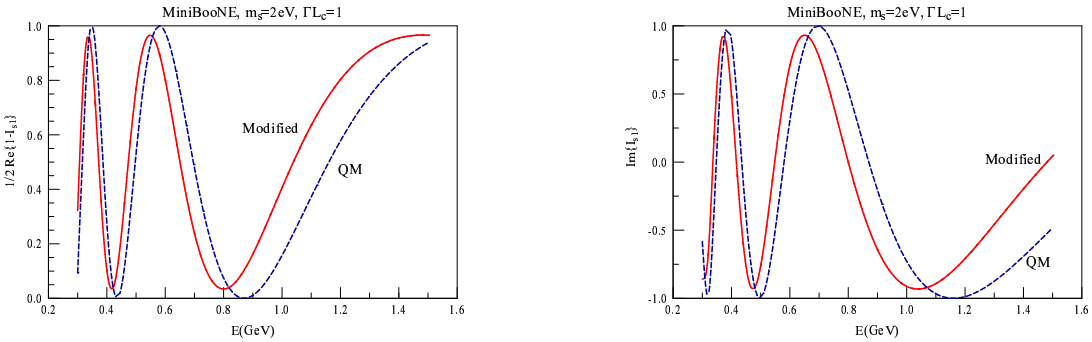


Figure 5: Same as Fig.(4) for MiniBooNE for $m_s = 2\text{eV}$.

Pions and Kaons (DAR): Recent proposals [81, 82] for high intensity sources to study sterile-active oscillations with pions and kaons (DAR) motivate a study of the decoherence

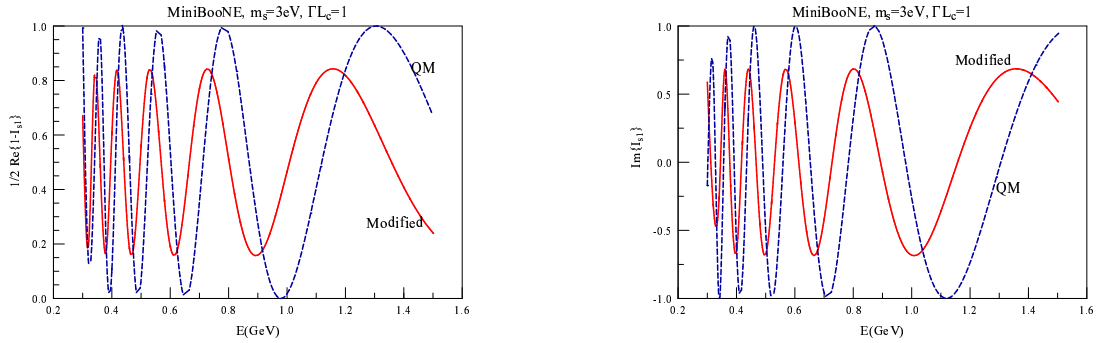


Figure 6: Same as Fig.(4) for MiniBooNE for $m_s = 3$eV.

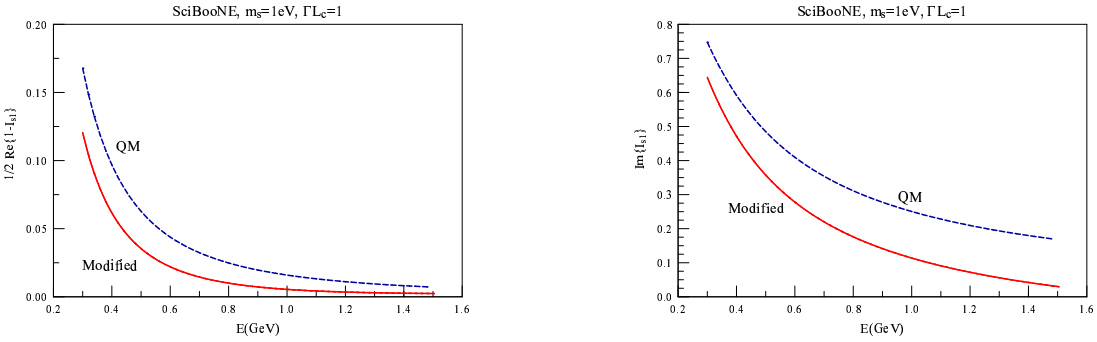


Figure 7: CP-even/odd parts of transition probability for SciBooNE parameters: $L = 100\text{m}, \Gamma_\pi L_c \simeq 1$ for $m_s = 1$eV. Solid line (modified) $\text{Re}[1 - I_{s1}]/2$ dashed line (Qm) is the quantum mechanical result $\sin^2[m^2_{\nu_s}/4E]$.

effects in these experiments. For (DAR) the energy is fixed at $E = E^*$ and presumably the baseline L is also fixed, we take $L = 30\,\text{m}$ as a middle-range indicative value for the purpose

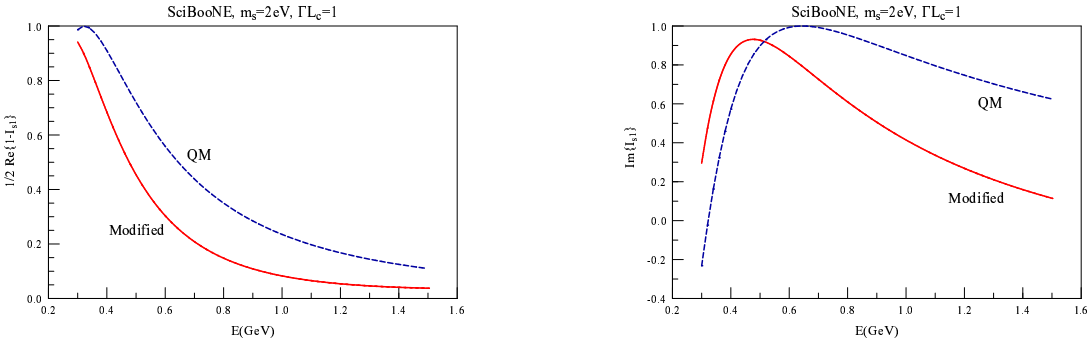


Figure 8: Same as Fig. (7) for SciBooNE for $m_s = 2$eV.

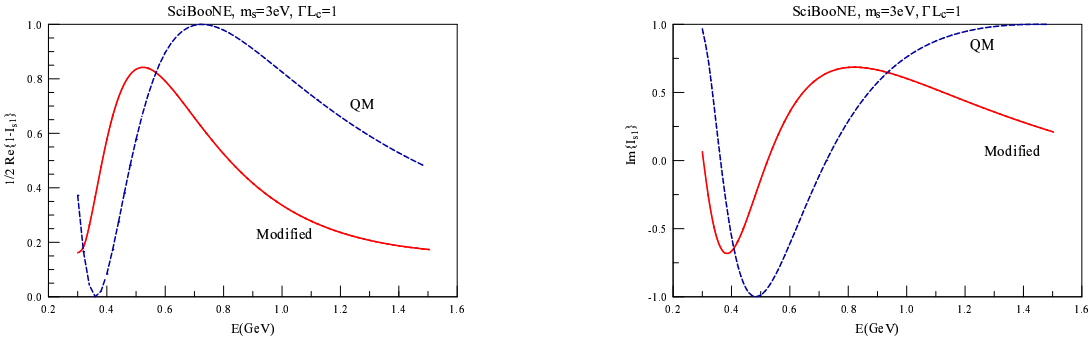


Figure 9: Same as Fig. (7) for SciBooNE for $m_s = 3$eV.

of our analysis, other values can be explored numerically. What is less clear is the value of the product $\Gamma_M L_c$ which will ultimately depend on the experimental design. Namely, the muons (or charged leptons in general) must be stopped at distances much less than the baseline and that $\Gamma L_c \ll 1$ in order for decoherence effects to be minimal. We study the

possibible ranges $\Gamma_M L_c \ll 1, \simeq 1, \gg 1$ respectively as a function of m_s. For $\pi - K$ (DAR) it follows that $E^*_\pi = 29.8\,\text{MeV}$; $E^*_K = 235.5\,\text{MeV}$ respectively for which we find the ratio (2.6.11) to be

$$\mathcal{R}_\pi(E^*_\pi) = \frac{2}{3}\left(\frac{m_s^2}{\text{eV}^2}\right) \quad ; \quad \mathcal{R}_K(E^*_K) = \frac{1}{25}\left(\frac{m_s^2}{\text{eV}^2}\right) \tag{2.6.15}$$

The comparison between the modified results (2.6.3,2.6.4) and the usual quantum mechanical results (2.6.6) are displayed in figs. (10-15) for π, K (DAR) for a baseline representative $L = 30\,\text{m}$.

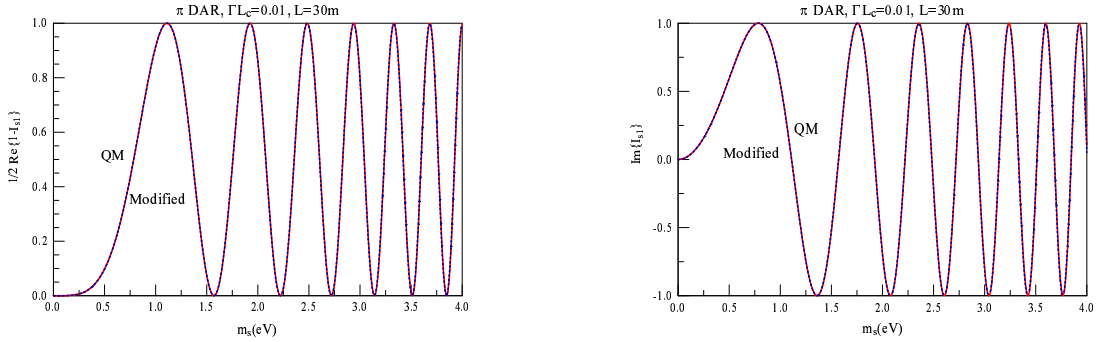


Figure 10: π (DAR) CP-even and CP-odd contributions for $\Gamma_\pi L_c = 0.01$ vs m_s. The solid line (modified) shows $\text{Re}[1 - I_{s1}]$ where $\text{Re}[I_{s1}]$ is given by (2.6.3,2.6.4), the dashed line is the quantum mechanical result (2.6.6) for $\delta m^2_{s1} = m^2_s$.

It is clear from this analysis, both for π, K (DAR), that decoherence effects are very small and the modified result is indistinguishable from the usual quantum mechanical results (2.6.6) whenever $\Gamma_M L_c \ll 1$ but become large for $\Gamma_M L_c \gtrsim 1$. The decay length for π, K are $l_\pi = 7.8\,\text{m}, l_K = 3.7\,\text{m}$ respectively; therefore, in order for the usual quantum mechanical results (2.6.6) to describe a correct fit to the experimental data, the design must ensure that charged leptons (mainly μ) *be stopped at distances* $L_c \ll l_\pi, l_K$ *respectively*, namely a few cm beyond the stopping target of the mesons.

Long baseline experiments: For long baseline experiments, the decoherence terms do

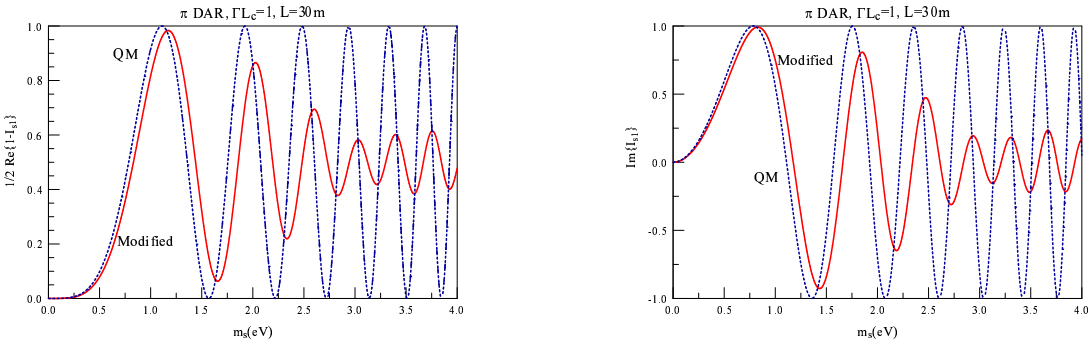


Figure 11: Same as fig. (10) with $\Gamma_\pi L_c = 1$.

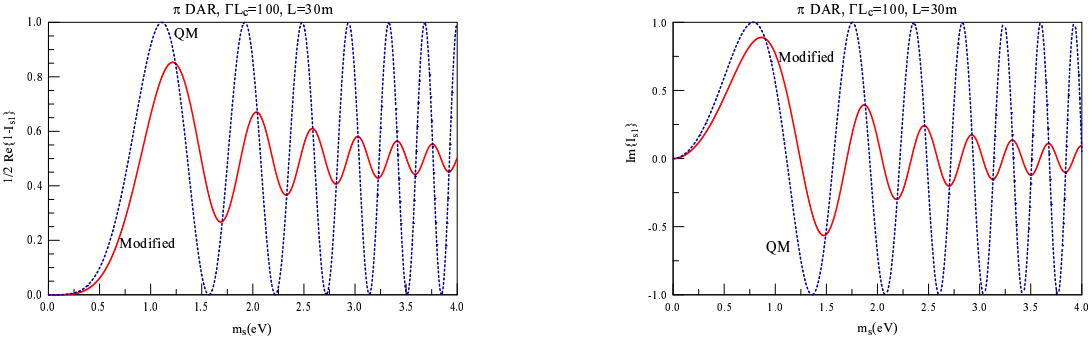


Figure 12: Same as fig. (10) with $\Gamma_\pi L_c = 100$.

not contribute. This is because these experiments study oscillations with $\delta m^2 \sim 10^{-3}\,\mathrm{eV}^2$ and $E \simeq$ few GeV for which $L \simeq 300 - 1000\,\mathrm{km}$, an example of such experiment is Minos in which pions produce neutrinos as in MiniBooNE/SciBoone. In these experiments, $\mathcal{R} \lesssim 10^{-3}$; $\Gamma_M L_c \lesssim 1$ so that $\Delta_{ij} L_c \ll 1$; therefore, decoherence effects are all but negligible

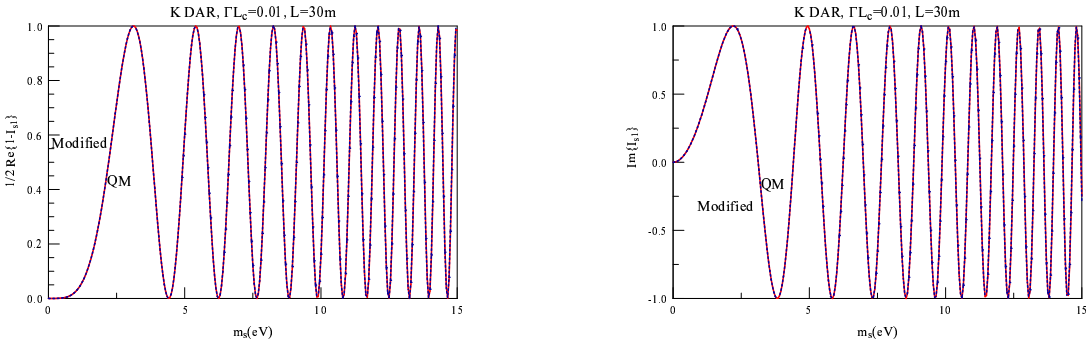


Figure 13: K (DAR) CP-even and CP-odd contributions for $\Gamma_K L_c = 0.01$ vs m_s. The solid line (modified) shows $\mathrm{Re}[1 - I_{s1}]$ where $\mathrm{Re}[I_{s1}]$ is given by (2.6.3,2.6.4), the dashed line is the quantum mechanical result (2.6.6) for $\delta m_{s1}^2 = m_s^2$.

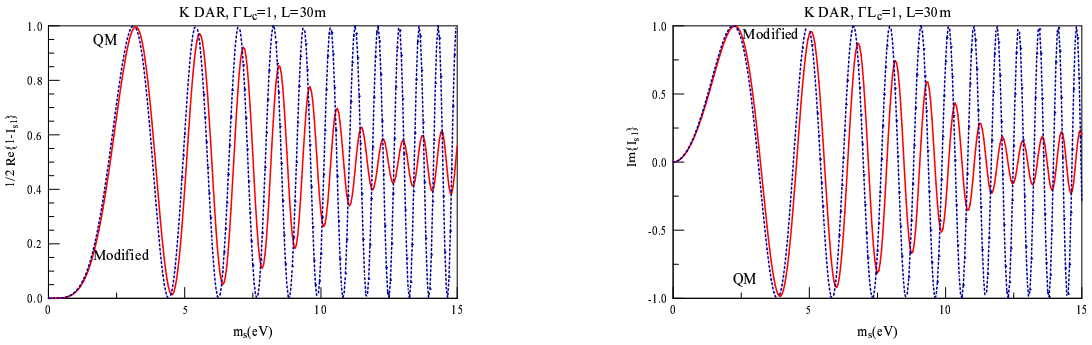


Figure 14: Same as fig. (13) with $\Gamma_K L_c = 1$.

generally for long baseline experiments.

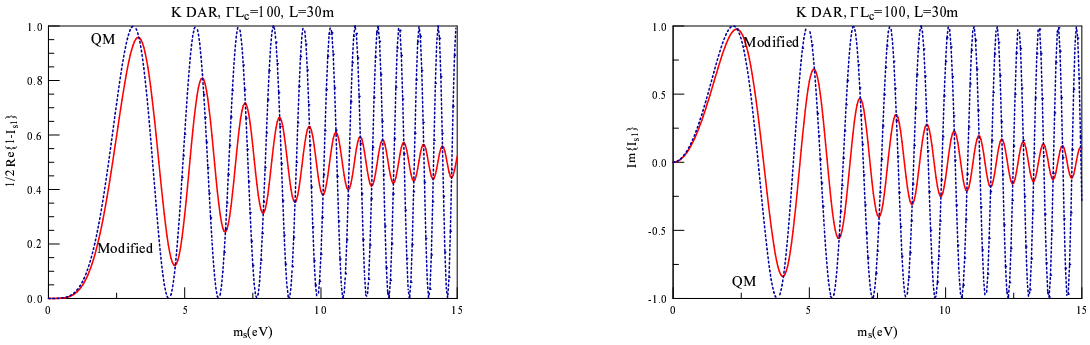


Figure 15: Same as fig. (13) with $\Gamma_K L_c = 100$.

2.6.1 Reactors vs. accelerator experiments

The suppression of the transition probabilities through the decoherence effects depend both on the lifetime of the parent particle and the stopping distance of the charged lepton which is produced along with the (anti) neutrino via the charged current interactions. This establishes a *fundamental difference between accelerator and reactor oscillation experiments*: whereas in accelerator experiments neutrinos are produced via the decay of short lived mesons with typical lifetimes $\simeq 10^{-8}$secs and widths $\simeq 10^{-8}$ eV, in reactors the (anti) neutrinos are produced via the β decay of long-lived unstable nuclei $^{235}\mathrm{U}\,, ^{238}\mathrm{U}\,, ^{239}\mathrm{Pu}\,, ^{241}\mathrm{Pu}$[47, 48] with typical lifetimes in the range between hundreds and thousands of years. Furthermore, in current short baseline accelerator experiments such as MiniBooNE/SciBooNE, pions decay in a decay pipe and muons are stopped at short distances beyond the decay pipe so that $\Gamma_\pi L_c \simeq 1$; in reactor experiments, muons are stopped in the reactor core on distance scales so that $\Gamma L_c\ ,\ \Delta_{ji} L_c \ll 1$. Our study above clearly shows that under this circumstance the modifications from decoherence are negligible and the transition probabilities are indistinguishable from the quantum mechanical result. Therefore, we conclude that the quantum mechanical fit to the oscillation probabilities in reactor experiments is always justified, whereas in accel-

erator experiments, decoherence effects both from the lifetime of the parent meson and the stopping length scale of the charged lepton partner are substantial for $m_s \gtrsim 1\,\text{eV}$ and a fit to the usual quantum mechanical transition probabilities both for CP-even/odd contributions may substantially *underestimate masses, mixing and CP-violating angles*.

2.6.2 Wave packets:

Our study is restricted to plane waves to exhibit the main results and conclusions in the clearest possible setting. As has been argued in the literature[109, 110, 111, 112, 113, 114, 115, 116], wave packet localization may be an important ingredient in the description of neutrino oscillations. The localization length both of the production and detection regions define momentum uncertainties that are important in the conceptual understanding of the interference phenomena. A wave packet description should also be implemented in the measurement or stopping (disentanglement) of the charged lepton, which we treated as an event sharp in space time at a time scale t_c and distance L_c, a wave packet treatment would smear these scales over a localization length scale of the wave packet, which is determined by the measurement process (or perhaps the mean free path of the charged lepton in the stopping material).

The typical analysis of neutrino oscillations in terms of (Gaussian) wave packets [109, 110, 111, 112, 113, 114, 115, 116] clarifies that neutrino wave-packets evolve semiclassically, the center moves as the front of a plane wave with the group velocity and is modulated by a Gaussian envelope which spreads via dispersion. Wave packets associated with the different mass eigenstates separate as they evolve with slightly different group velocities and, when their separation becomes of the order of or larger than the width of the wave packet, the overlap vanishes and oscillations are suppressed $\propto e^{-(L/L_{coh})^2}$, where $L_{coh} \simeq \sigma\, E_\nu^2/\delta m^2$ and σ is the spatial localization scale of the wave packet. This suppression becomes important when $L_{coh} \lesssim L$, which for $\delta m^2 \simeq \text{eV}^2, E \gtrsim 30\,\text{MeV}, L \sim 100 - 600\,\text{m}$ implies $\sigma \lesssim 1 - 5 \times 10^{-13}\,\text{m}$ which, while much bigger than nuclear dimensions, is much smaller than atomic scales. If a firm assessment confirms that neutrino wavepackets are produced with such localization length or smaller, then this decoherence effect must be introduced in the

oscillation probability.

As discussed in [109] the wave packet description also features another source of decoherence in the *localization term*, which suppresses coherence when $\sigma > L_{osc} \sim E/\delta m^2$ which is unlikely to be relevant in short baseline accelerator experiments. A complementary interpretation of decoherence for $\Gamma_M \lesssim \delta m^2/2E$ in terms of wave packets is discussed in ref.[109]: *if* a neutrino wavepacket produced by the decay of a parent particle of width Γ_M is assigned a localization length, $\sigma \simeq 1/\Gamma_M$, then the condition for decoherence from the *localization term*, $\sigma \simeq L_{osc}$, becomes equivalent to $\Gamma_M \simeq \delta m^2/2E$ which is recognized as $\mathcal{R} \simeq 1$ in our discussion. Although we do not see an obvious relation between the results obtained above with the non-perturbative field theoretical Wigner-Weisskopf method and the interpretation of a wavepacket with localization length $1/\Gamma_M$, our results are certainly in agreement with this interpretation; however, we emphasize that the analysis above also reveals another scale that is important for decoherence, namely L_c, which is the length scale at which the charged lepton that is emitted along with the neutrino is observed or absorbed. As pointed out above there are *two* important dimensionless quantities that determine decoherence in the plane wave limit: $\mathcal{R}$ and $\Gamma_M L_c$.

2.7 CONCLUSIONS AND FURTHER QUESTIONS

Motivated by the cosmological importance of new generations of heavier sterile neutrinos and recent proposal for high intensity sources, this article focuses on two different aspects related to the search of sterile neutrinos: 1) a proposal to search for heavy ($\simeq$ MeV-range) sterile neutrinos by studying the production of *negative helicity* charged leptons in π^-, K^- decay at rest (or positive helicity in the decay of π^+, K^+) as a complement to the search for monochromatic lines in the muon (or electron) spectrum, and 2) an assessment of the impact of decoherence effects from the lifetime of the parent meson and the stopping distance scale of the charged lepton on the experimental fits for sterile neutrinos masses, mixing and CP-violating angles in short baseline experiments.

Massive sterile (anti) neutrinos produced in π^-, K^- decay at rest (DAR) lead to a *negative* helicity (positive if the decay is π^+, K^+) component of the charged lepton produced in the decay. For searches of heavy sterile neutrinos from π^-, K^- decay at rest, we obtain the branching ratio for charged leptons to be produced with *negative helicity* (or positive helicity for the decay of π^+, K^+). This branching ratio determines the abundance of the negative helicity states in the production process and we suggest that a Stern-Gerlach type filter with a magnetic field with a gradient along the direction of the collimated charged lepton beam emitted back to back with the (anti) neutrinos allows to spatially separate the different helicity components. A combined measurement of the monochromatic line for the charged lepton and the ratio of abundances of the spatial domains yield simultaneous information on the mass and the absolute value of the mixing matrix element. This setup is most sensitive for *heavy* sterile neutrinos with mass m_s in the MeV range. The ratio of abundances between the negative and positive helicity states is determined by the branching ratio (2.2.15), shown in fig. (1) (divided by $|U_{ls}|^2$), which, in combination with the search for monochromatic lines allows, to extract both the mass and the element of the mixing matrix $|U_{ls}|^2$ by fitting both the energy and abundance with the branching ratios.

Upper bounds on the sterile-active mixing matrix elements from previous experimental searches allow us to estimate the upper bounds for the branching ratios for the different processes, these are given by

$$Br^{--}_{\pi\to\mu\,\bar{\nu_s}} \lesssim 10^{-8} - 10^{-7} \;\; ; \;\; 3\,\text{MeV} \lesssim m_s \lesssim 33\,\text{MeV} \tag{2.7.1}$$

$$Br^{--}_{\pi\to e\,\bar{\nu_s}} \lesssim 10^{-8} - 10^{-6} \;\; ; \;\; 3\,\text{MeV} \lesssim m_s \lesssim 135\,\text{MeV} \tag{2.7.2}$$

with the electron channel providing the largest window of opportunity because of the larger phase space. For K-(DAR), we find

$$Br^{--}_{K\to\mu,e\,\bar{\nu_s}} \lesssim 10^{-9} - 10^{-6} \text{ for } \begin{cases} 4\,\text{MeV} \lesssim m_s \lesssim 360\,\text{MeV} \;\; (\mu - \text{channel}) \\ 4\,\text{MeV} \lesssim m_s \lesssim 414\,\text{MeV} \;\; (e - \text{channel}) \end{cases} . \tag{2.7.3}$$

These upper bounds estimates suggest that these searches could be implemented in the next generation of high intensity experiments.

Short baseline experiments target new generation of sterile neutrinos in the mass range $\simeq$ eV as suggested by the LSND, MiniBooNE results and reactor anomalies. In current accelerator experiments, (anti) neutrinos are produced from the decay of pions or kaons either in flight, as in MiniBooNE/SciBooNE, or at rest, as recent proposals suggest. We recognized *two* sources of decoherence that impact the interpretation of the data and experimental fits to extract masses, mixing and CP-violating angles: a) the width of the parent meson Γ_M introduces an energy (or time) uncertainty and b) the stopping distance L_c of the charged lepton that is produced in a quantum entangled state with the (anti) neutrino, decoherence effects are encoded in two different dimensionless quantities,

$$\mathcal{R}_{ij}(E) = \frac{\delta m^2_{ij}}{2E\Gamma_M} \;\; ; \;\; \Gamma_M L_c \, . \tag{2.7.4}$$

The usual quantum mechanical formula for the oscillation probabilities are modified as follows:

$$e^{i\frac{\delta m^2_{ij}}{2E}L} \to e^{i\frac{\delta m^2_{ij}}{2E}L}\left[\frac{1-e^{-\Gamma_M(p)L_c\left(1+i\mathcal{R}_{ij}\right)}}{1-e^{-\Gamma_M(p)L_c}}\right]\left[\frac{1-i\mathcal{R}_{ij}}{1+\mathcal{R}^2_{ij}}\right] \tag{2.7.5}$$

We study the impact of the decoherence effects both for Dirac and Majorana neutrinos, addressing in particular CP-violating effects as well as $\nu \to \overline{\nu}$ oscillations and $|\Delta L| = 2$ transitions in the case of Majorana neutrinos. In all cases, we find that, for $\mathcal{R}_{ij}, \Gamma_M L_c \gtrsim 1$, the oscillation probabilities are suppressed and the oscillatory functions feature energy-dependent phase-shifts that results in an overall off-set that impacts the determination of the mass. If these decoherence effects are neglected in the experimental analysis and the data are fit with the usual quantum mechanical oscillation probabilities the masses, mixing and CP-violating angles are *underestimated*.

In particular, on MiniBooNE/SciBooNE, for example, neutrinos are produced from pion decay for which we find $\mathcal{R} \simeq 1/3\left(\delta m^2/\mathrm{eV}^2\right)$ and $\Gamma_M L_c \simeq 1$, with one sterile neutrino with $m_s \sim 3\,\mathrm{eV}$, fitting with two-generation mixing underestimates $\sin^2(2\theta)$ and δm^2 by nearly 15%. Similar underestimates follow for CP-violating angles and $|\Delta L| = 2$ processes in $3+2$ schemes.

We also conclude that reactor and (current) accelerator experiments are fundamentally different in that the lifetime of the decaying parent particles in reactor experiments is hundreds to thousands of years, compared to pion or kaon lifetimes, and charged leptons (muons) are stopped within the core so that for reactors $\Gamma L_c, \Delta_{ij} L_c \ll 1$ and decoherence effects are all but negligible, unlike the situation for example for MiniBooNE/SciBooNE. We also suggest that next generation of high intensity experiments in which (anti) neutrinos are produced from π, K (DAR), decoherence effects may be suppressed considerably by designing the experiment so that the charged leptons produced with the neutrinos (mainly muons) are stopped on distances much smaller than the decay length of the mesons, in which case the usual quantum mechanical oscillation probabilities furnish an accurate description of mixing and oscillations.

2.8 APPENDICES

2.8.1 Quantization: Mesons, Dirac and Majorana neutrinos

We quantize the (pseudo) scalar and fermion fields in a quantization volume V. The charged (complex) (pseudo) scalar field is as usual

$$M(\vec{x},t) = \sum_{\vec{p}} \frac{1}{\sqrt{2VE_M(p)}} \left[\hat{\mathcal{A}}_{\vec{p}}\, e^{-iE_M(p)\,t} + \hat{\mathcal{B}}^{\dagger}_{-\vec{p}}\, e^{iE_M(p)\,t} \right] e^{i\vec{p}\cdot\vec{x}} \tag{2.8.1}$$

where $E_p^M = \sqrt{p^2 + m_M^2}$ with m_M the mass of the corresponding meson. It follows that

$$J^M_\mu(\vec{x},t) = \sum_{\vec{p}} \frac{p_\mu}{\sqrt{2VE_M(p)}} \left[\hat{\mathcal{A}}_{\vec{p}}\, e^{-iE_M(p)\,t} - \hat{\mathcal{B}}^{\dagger}_{-\vec{p}}\, e^{iE_M(p)\,t} \right] e^{i\vec{p}\cdot\vec{x}} \tag{2.8.2}$$

It proves convenient to introduce the combinations

$$M^+(\vec{p},t) \pm M^-(\vec{p},t) = \left(\hat{\mathcal{A}}_{\vec{p}}\, e^{-iE_M(p)\,t} \right) \pm \left(\hat{\mathcal{B}}^{\dagger}_{-\vec{p}}\, e^{iE_M(p)t} \right) . \tag{2.8.3}$$

For Fermi fields we work in the chiral representation,

$$\gamma^0 = \begin{bmatrix} 0 & -\mathbb{1} \\ -\mathbb{1} & 0 \end{bmatrix} \;\; ; \;\; \gamma^i = \begin{bmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{bmatrix} \;\; ; \;\; \gamma^5 = \begin{bmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{bmatrix} \tag{2.8.4}$$

and for a generic Fermion f, either for charged lepton or Dirac neutrinos of mass m_f, we write

$$\Psi(\vec{x},t) = \sum_{h=\pm} \sum_{\vec{k}} \frac{\psi(\vec{k},h,t)}{\sqrt{2VE_f(k)}} \;\; e^{i\vec{k}\cdot\vec{x}} \tag{2.8.5}$$

For *Dirac* fermions of mass m_f

$$\psi(\vec{k},h,t) = \left[\hat{b}_{\vec{k},h} \mathcal{U}_h(\vec{k})\, e^{-iE_f(k)\,t} + \hat{d}^\dagger_{-\vec{k},h} \mathcal{V}_h(\vec{k})\, e^{iE_f(k)\,t} \right] \tag{2.8.6}$$

where $E_f(k) = \sqrt{k^2 + m_f^2}$ and the spinors $\mathcal{U}_h, \mathcal{V}_h$ are eigenstates of helicity with eigenvalue $h = \pm 1$, these are given by

$$\mathcal{U}_+(\vec{k}) \;\; = \;\; N_f \begin{pmatrix} v_+(\vec{k}) \\ -\varepsilon(k)\, v_+(\vec{k}) \end{pmatrix} \;\; ; \;\; \mathcal{U}_-(\vec{k}) = N_f \begin{pmatrix} -\varepsilon(k)\, v_-(\vec{k}) \\ v_-(\vec{k}) \end{pmatrix} \tag{2.8.7}$$

$$\mathcal{V}_+(\vec{k}) \;\; = \;\; N_f \begin{pmatrix} \varepsilon(k)\, v_+(\vec{k}) \\ v_+(\vec{k}) \end{pmatrix} \;\; ; \;\; \mathcal{V}_-(\vec{k}) = N_f \begin{pmatrix} v_-(\vec{k}) \\ \varepsilon(k)\, v_-(\vec{k}) \end{pmatrix} \tag{2.8.8}$$

where

$$N_f = \sqrt{E_f(k) + k} \;\; ; \;\; \varepsilon(k) = \frac{m_f}{E_f(k) + k} \tag{2.8.9}$$

and $v_\pm(\vec{k})$ are helicity eigenstates Weyl spinors:

$$v_+(\vec{k}) = \begin{pmatrix} \cos\frac{\theta}{2} \\ \sin\frac{\theta}{2}\, e^{i\phi} \end{pmatrix} \;\; ; \;\; v_-(\vec{k}) = \begin{pmatrix} -\sin\frac{\theta}{2}\, e^{-i\phi} \\ \cos\frac{\theta}{2} \end{pmatrix} \tag{2.8.10}$$

where

$$\vec{k} = k \left(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta \right). \tag{2.8.11}$$

A useful representation is

$$v_+(\vec{k}) = \frac{\left(1 + \vec{\sigma}\cdot\hat{\vec{k}}\right)}{\sqrt{2(1+\cos\theta)}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \;\; ; \;\; v_-(\vec{k}) = \frac{\left(1 - \vec{\sigma}\cdot\hat{\vec{k}}\right)}{\sqrt{2(1+\cos\theta)}} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{2.8.12}$$

The Weyl spinors (2.8.10) satisfy

$$v_h^\dagger(\vec{k}) \cdot v_{h'}(\vec{k}) = \delta_{h,h'} \tag{2.8.13}$$

Majorana fields are charge self-conjugate and generally obey

$$\psi^c = i\gamma^2\,\psi^* = e^{i\xi}\,\psi \tag{2.8.14}$$

with ξ an arbitrary (real) phase, which we choose $\xi = 0$. In the chiral representation (2.8.4) writing

$$\psi = \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} \tag{2.8.15}$$

it follows that

$$\psi^c = \begin{pmatrix} i\sigma^2\,\psi_L^* \\ -i\sigma^2\,\psi_R^* \end{pmatrix} \tag{2.8.16}$$

Therefore, a Majorana field is obtained by combining the positive frequency component with its charge conjugate as the negative frequency, namely

$$\chi(\vec{x},t) = \sum_{h=\pm}\sum_{\vec{k}} \frac{1}{\sqrt{2VE_f(k)}} \left[\hat{b}_{\vec{k},h}\mathcal{U}_h(\vec{k})\,e^{-i(E_f(k)\,t-\vec{k}\cdot\vec{x})} + \hat{b}^\dagger_{\vec{k},h}\mathcal{U}^c_h(\vec{k})\,e^{i(E_f(k)\,t-\vec{k}\cdot\vec{x})} \right] \tag{2.8.17}$$

where

$$\mathcal{U}^c_+(\vec{k}) = N_f \begin{pmatrix} \varepsilon(k)\,v_-(\vec{k}) \\ v_-(\vec{k}) \end{pmatrix} \;\; ; \;\; \mathcal{U}^c_-(\vec{k}) = N_f \begin{pmatrix} v_+(\vec{k}) \\ \varepsilon(k)\,v_+(\vec{k}) \end{pmatrix} \tag{2.8.18}$$

and we have used the property

$$\big(i\sigma^2\big)v^*_+(\vec{k}) = -v_-(\vec{k}) \;\; ; \;\; \big(i\sigma^2\big)v^*_-(\vec{k}) = v_+(\vec{k})\,. \tag{2.8.19}$$

In particular the negative chirality component of the Majorana neutrino is

$$\begin{aligned} \chi_L(\vec{x},t) = \quad & \frac{1}{\sqrt{V}}\sum_{\vec{k}} \left[\frac{E_f(k)+k}{2E_f(k)}\right]^{\frac{1}{2}} \left[\left(-\hat{b}_{\vec{k},+}\,\varepsilon(k)\,v_+(\vec{k}) + \hat{b}_{\vec{k},-}\,v_-(\vec{k}) \right) e^{-i(E_f(k)\,t-\vec{k}\cdot\vec{x})} \right. \\ + \quad & \left. \left(\hat{b}^\dagger_{\vec{k},+}\,v_-(\vec{k}) + \hat{b}^\dagger_{\vec{k},-}\varepsilon(k)\,v_+(\vec{k}) \right) e^{i(E_f(k)\,t-\vec{k}\cdot\vec{x})} \right] \end{aligned} \tag{2.8.20}$$

From the representation (2.8.12), it follows that

$$v_h^\dagger(-\vec{k})\cdot v_h(\vec{k}) = 0 \quad ; \quad h = \pm\,. \tag{2.8.21}$$

It is straightforward to confirm that the Hamiltonian for the Majorana fields

$$\frac{1}{2}\int d^3x \chi^\dagger(\vec{x},t)\Big[-i\vec{\alpha}\cdot\vec{\nabla}+\beta m_f\Big]\chi(\vec{x},t) = \sum_{k,h} E_f(k)\,\hat{b}^\dagger_{\vec{k},h}\,\hat{b}_{\vec{k},h} \tag{2.8.22}$$

where the zero point energy has been subtracted.

2.8.2 Wigner-Weisskopf method for $M \to l\overline{\nu}$:

The purpose of this appendix is to provide technical details of the Wigner-Weisskopf approximation as applied to the $M \to l\bar{\nu}$ process. For a more extended discussion see refs.[85, 175].

The total Hamiltonian is given by $H = H_0 + H_i$, where H_0 is the free Hamiltonian and H_i is the interaction part. The time evolution of a state in the interaction picture is given by

$$i\frac{d}{dt}|\Psi(t)\rangle_I = \hat{H}_I|\Psi(t)\rangle_I \tag{2.8.23}$$

where $\hat{H}_I(t) = e^{i\hat{H}_0 t}\hat{H}_i(t)e^{-i\hat{H}_0 t}$. The formal solution of (2.8.23) is given by

$$|\Psi(t)\rangle_I = \hat{U}(t,t_o)|\Psi(t_o)\rangle_I \tag{2.8.24}$$

where $\hat{U}(t,t_o) = T(e^{-i\int_{t_o}^{t}\hat{H}_I(t')dt'})$. Expanding the state $|\Psi(t)\rangle_I$ in the basis of eigenstates of H_0 we have

$$|\Psi(t)\rangle_I = \sum_n C_n(t)|n\rangle \tag{2.8.25}$$

where $\hat{H}_0|n\rangle = E_n|n\rangle$. It is straightforward to show that $\sum_n |C_n(t)|^2 = const$ which is a consequence of unitary time evolution.

Now consider the initial state at time $t = 0$ to be one meson state of definite momentum, namely

$$|\Psi(t=0)\rangle_I = |M\rangle_{\vec{p}} = \sum_n C_n(t=0)|n\rangle \tag{2.8.26}$$

which gives the initial condition $C_n(t=0) = \delta_{n,M_{\vec{p}}}$.

From eq.(2.8.23), upon expanding in basis states, it follows that

$$\frac{d}{dt}C_n(t) = -i\sum_m \langle n|H_I(t)|i\rangle C_m(t) \tag{2.8.27}$$

The interaction Hamiltonian (2.2.1) connects the initial meson state, $|M_{\vec{p}}\rangle$ to leptonic / neutrino states, $\{|l\rangle \otimes |\bar{\nu}\rangle\}$. These states in turn are coupled back to $|M_{\vec{p}}\rangle$ via H_I, but also to other multiparticle states which describe processes that are higher order in perturbation theory. However, we will only be considering states connected to $|M_{\vec{p}}\rangle$ via first order in perturbation theory. The case that will be of interest to us will be $M \to l\bar{\nu}$ and is shown in Figure (16).

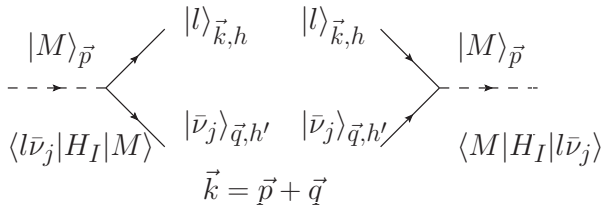


Figure 16: Transitions $|M\rangle \to |l\rangle|\bar{\nu}_j\rangle$

Considering the set of equations for these states, we obtain

$$\frac{d}{dt}\,C_M(t) = -i\sum_\kappa \langle M|H_I(t)|\kappa\rangle C_\kappa(t) \tag{2.8.28}$$

$$\frac{d}{dt}\,C_\kappa(t) = -i\langle \kappa|H_I(t)|M\rangle C_M(t) \tag{2.8.29}$$

where $|\kappa\rangle$ is the intermediate state, $|l_\alpha(\vec{k},h)\rangle|\bar{\nu}_\alpha(\vec{q},h')\rangle$. Using the initial conditions for $t = 0$, one obtains

$$C_\kappa(t) = -i\int_0^t dt'\langle \kappa|H_I(t')|M\rangle C_M(t')\,, \tag{2.8.30}$$

which when inserted into (2.8.28) leads to

$$\frac{d}{dt}C_M(t) = -\int_0^t dt'\sum_\kappa \langle M|H_I(t)|\kappa\rangle\langle \kappa|H_I(t')|M\rangle C_M(t') = -\int_0^t dt'\Sigma_M(t-t')C_M(t') \tag{2.8.31}$$

Where the meson self energy has been introduced

$$\Sigma_M(t-t') \equiv \sum_\kappa \langle M|H_I(t)|\kappa\rangle\langle\kappa|H_I(t')|M\rangle = \sum_\kappa |\langle M|\hat{H}_I(0)|\kappa\rangle|^2 e^{i(E_M-E_\kappa)(t-t')} \tag{2.8.32}$$

This self-energy is recognized as the one-loop retarded self energy with the $|l\rangle|\overline{\nu}\rangle$ intermediate state.

Solving eq.(2.8.31) produces a solution for the time evolution of the meson amplitude. We can use the solution for $C_M(t)$ to obtain an expression for the amplitudes $C_\kappa(t)$ which allows for computation of the probability of occupying a particular state at any given time. We may solve eq.((2.8.31)) either via Laplace transform, or in the case of weak coupling, a derivative expansion which yields the same result at long times ($t \gg 1/m_M$). Here, we follow the latter method which is the original Wigner-Weisskopf approximation.

We begin by defining the quantity

$$W_0(t,t') = \int_0^{t'} dt'' \Sigma_M(t-t'') \tag{2.8.33}$$

so that

$$\frac{d}{dt'} W_0(t,t') = \Sigma_M(t-t') \quad , \quad W_0(t,0) = 0 \tag{2.8.34}$$

Integrating eq.(2.8.31) by parts yields

$$\frac{d}{dt} C_M(t) = -\int_0^t dt' \Sigma_M(t-t') C_M(t') = -W_0(t,t) C_M(t) + \int_0^t dt' W_0(t,t') \frac{d}{dt'} C_M(t') \tag{2.8.35}$$

The first is term second order in H_I whereas the second term is of fourth order in H_I and will be neglected. This approximation is equivalent to the Dyson resummation of the one-loop self energy diagrams. Thus to leading order, eq.(2.8.31) becomes

$$\frac{d}{dt} C_M(t) + W_0(t,t) C_M(t) = 0\,, \tag{2.8.36}$$

where

$$W_0(t,t) = \int_0^t dt' \Sigma_M(t-t') = \int_0^t dt' \sum_\kappa |\langle M|\hat{H}_I(0)|\kappa\rangle|^2 e^{i(E_M - E_\kappa)(t-t')} \tag{2.8.37}$$

Inserting a convergence factor and taking the limit $t \to \infty$ consistently with the Wigner-Weisskopf approximation, we find[6]

$$W_0(t,t) = \lim_{\epsilon \to 0^+} i \frac{\sum_\kappa |\langle M|\hat{H}_I(0)|\kappa\rangle|^2}{E_M - E_\kappa + i\epsilon} = i\Delta E_M + \frac{\Gamma_M}{2} \tag{2.8.38}$$

where

$$\Delta E_M \equiv \mathcal{P} \sum_\kappa \frac{|\langle M|\hat{H}_I(0)|\kappa\rangle|^2}{E_M - E_\kappa}\,, \tag{2.8.39}$$

is the second order shift in the energy which will be absorbed into a renormalized meson energy and

$$\Gamma_M \equiv 2\pi \sum_\kappa |\langle M|\hat{H}_I(0)|\kappa\rangle|^2 \delta(E_M - E_\kappa) \tag{2.8.40}$$

is the decay width as per Fermi's Golden rule. Therefore in this approximation, we arrive at

$$C_M(t) = e^{-i\Delta E_M t} e^{-\frac{\Gamma_M}{2}t}\,. \tag{2.8.41}$$

Inserting this result into eq. (2.8.30) leads to

$$\begin{aligned} C_\kappa(t) &= -i\langle\kappa|H_I(0)|M\rangle \int_0^t dt' e^{-i(E_M + \Delta E_M - E_\kappa - i\frac{\Gamma_M}{2})t'} \\ &= -\langle\kappa|H_I(0)|M\rangle \left[\frac{1 - e^{-i(E_M + \Delta E_M - E_\kappa - i\frac{\Gamma_M}{2})t}}{E_A + \Delta E_M - E_\kappa - i\frac{\Gamma_M}{2}}\right] \end{aligned} \tag{2.8.42}$$

[6]The long time limit in the Wigner-Weisskopf approximation is equivalent to the Breit-Wigner approximation of a resonant propagator[175].

Defining the renormalized energy of the single particle meson state as $E_M^r = E_M + \Delta E_M$ and passing to the Schroedinger picture $|M(t)\rangle_S = e^{-i\hat{H}_0 t}|M(t)\rangle_I$, we find that

$$\begin{aligned}|M^-_{\vec{p}}(t)\rangle_S &= e^{-i\hat{H}_0 t}\left[C_M(t)|M\rangle + \sum_\kappa C_\kappa(t)|\kappa\rangle\right] \\ &= e^{-iE_M^r t}e^{-\frac{\Gamma_M}{2}t}|M^-_{\vec{p}}(0)\rangle - \sum_\kappa \langle\kappa|H_I(0)|M^-_{\vec{p}}\rangle\left[\frac{1-e^{-i(E_M^r - E_\kappa - i\frac{\Gamma_M}{2})t}}{E_M^r - E_\kappa - i\frac{\Gamma_M}{2}}\right]e^{-iE_\kappa t}|\kappa\rangle\end{aligned} \tag{2.8.43}$$

The interaction Hamiltonian for $M \to l_\alpha \bar{\nu}_\alpha$ is given by eqn. (2.2.5) and the quantization from Appendix A leads to the matrix element

$$\langle l^-_\alpha \overline{\nu}|H_I(0)|M^-_{\vec{p}}\rangle = \frac{F_M}{\sqrt{V}}\sum_j U_{\alpha j}\frac{\overline{\mathcal{U}}_{\alpha,h}(\vec{k})\gamma^\mu \mathbb{L}\mathcal{V}_{j,h'}(\vec{q})p_\mu}{\sqrt{8E_M(p)E_\alpha(k)E_j(q)}} \quad ; \quad \vec{k} = \vec{p} + \vec{q} \tag{2.8.44}$$

which yields our final result for the entangled quantum state resulting from meson decay

$$\begin{aligned}|M^-_{\vec{p}}(t))\rangle = e^{-iE_M(p)t}e^{-\Gamma_M(p)\frac{t}{2}}|M^-_{\vec{p}}(0)\rangle - F_M \sum_{\vec{q},\alpha j,h,h'} U_{\alpha j}\frac{\overline{\mathcal{U}}_{\alpha,h}(\vec{k})\gamma^\mu \mathbb{L}\mathcal{V}_{j,h'}(\vec{q})p_\mu}{\sqrt{8VE_M(p)E_\alpha(k)E_j(q)}}\times \\ \left[\frac{1-e^{-i(E_M^r(p)-E_\alpha(k)-E_j(q)-i\frac{\Gamma_M}{2})t}}{E_M^r(p)-E_\alpha(k)-E_j(q)-i\frac{\Gamma_M}{2}}\right]e^{-i(E_\alpha(k)+E_j(q))t}|l^-_\alpha(h,\vec{k})\rangle|\bar{\nu}_j(h',-\vec{q})\rangle\end{aligned} \tag{2.8.45}$$

2.8.3 On the normalization (2.3.6):

The normalization of the disentangled neutrino state (2.3.6) has another important interpretation, it is recognized as the number density of charged leptons produced from the decay of the meson. To see this, consider the expansion of the Dirac field for the charged lepton as in eqn. (2.8.6) where $\hat{b}^\dagger_{\vec{k},h_i}$ creates a charged lepton l^- with momentum $\vec{k}$ and helicity h_i. The number operator for *particles* is $\hat{n}_{\vec{k},h_i} = \hat{b}^\dagger_{\vec{k},h_i}\hat{b}_{\vec{k},h_i}$ and its expectation value in the *full meson state* (2.3.1) is given by

$$n^l_{\vec{k},h_i} \equiv \langle M^-_{\vec{p}}(t))|\hat{n}_{\vec{k},h_i}|M^-_{\vec{p}}(t))\rangle = \sum_{j,h'}\frac{\left|U_{\alpha j}\right|^2\left|\mathcal{M}^{\mathcal{P}}_{\alpha j}(\vec{k},\vec{q},h_i,h')\right|^2}{8VE_M(pE_j(q)E_\alpha(k)}\left|\mathcal{F}_{\alpha j}[\vec{k},\vec{q};t_c]\right|^2(\vec{q};h_i)\,, \tag{2.8.46}$$

which is recognized as the normalization (2.3.6), namely

$$\mathcal{N}_\nu(\vec{q};h_i) = n^l_{\vec{k},h_i}\,. \tag{2.8.47}$$

From the definition of the partial width $\Gamma_{M^-\to l^-_\alpha \overline{\nu}_j}(p,h_i,h')$ of meson decay into a lepton α of helicity h and neutrino eigenstate ν_i of helicity h'

$$\Gamma_{M^-\to l^-_\alpha \overline{\nu}_j}(p,h_i,h') = \frac{1}{2E_M(p)}\int \frac{d^3q}{(2\pi)^3}\,\frac{\left|\mathcal{M}^{\mathcal{P}}_{\alpha j}(\vec{k},\vec{q},h_i,h')\right|^2}{2E_j(q)2E_\alpha(k)}\,2\pi\,\delta\big(E_M(p)-E_\alpha(|\vec{p}-\vec{q}|)-E_j(q)\big) \tag{2.8.48}$$

and the total decay width

$$\Gamma_M(p) = \sum_{j,h'}\left|U_{\alpha j}\right|^2\Gamma_{M^-\to l^-_\alpha \overline{\nu}_j}(p,h_i,h')\,, \tag{2.8.49}$$

it follows that the total number of charged leptons produced at time t_c is given by

$$V\sum_{h_i}\int\frac{d^3q}{(2\pi)^3}\,n^l_{\vec{k},h_i} = V\sum_{h_i}\int\frac{d^3q}{(2\pi)^3}\,\mathcal{N}_\nu(q,h_i) = \left[1-e^{-\Gamma_M(p)t_c}\right] \tag{2.8.50}$$

3.0 CHARGED LEPTON MIXING VIA HEAVY STERILE NEUTRINOS

Based on: (ref. [339])

L. Lello, D. Boyanovsky, Nuclear Physics B, Volume 880, March 2014, Pages 109-133

3.1 INTRODUCTION

Neutrino masses, mixing and oscillations are the clearest evidence yet of physics beyond the standard model [26, 27, 28, 29]. Oscillations among three "active" neutrinos with $\delta m^2 = 10^{-4} - 10^{-3}\,\mathrm{eV}^2$ for atmospheric and solar oscillations respectively have been firmly confirmed experimentally (see the reviews[33]-[118]).

However, several experimental hints have been accumulating that cannot be interpreted as mixing and oscillations among three "active" neutrinos with $\delta m^2 \simeq 10^{-4} - 10^{-3}$. Early results from the LSND experiment[43] have recently been confirmed by MiniBooNE running in antineutrino mode[44] both suggesting the possibility of new "sterile" neutrinos with $\delta m^2 \sim \mathrm{eV}^2$. The latest report from the MiniBooNE collaboration[45] on the combined $\nu_\mu \to \nu_e$ and $\overline{\nu}_\mu \to \overline{\nu}_e$ *appearance* data is consistent with neutrino oscillations with $0.01 < \Delta m^2 < 1.0\,\mathrm{eV}^2$. This is consistent with the evidence from LSND antineutrino oscillations[43], which bolsters the case for the existence of sterile neutrinos; however, combined MiniBooNE/SciBooNE analysis[46] of the $\overline{\nu}_\mu$ *disappearance* data are consistent with *no short baseline disappearance* of $\overline{\nu}_\mu$. Recently, a re-examination of the antineutrino flux[47] in anticipation of the Double Chooz reactor experiment resulted in a small increase in the

flux of about 3.5% for reactor experiments leading to a larger deficit of 5.7% suggesting a *reactor anomaly*[48]. If this deficit is the result of neutrino mixing and oscillation with baselines $L \lesssim 10 - 100\,\mathrm{m}$, it requires the existence of at least one sterile neutrino with $\delta m^2 \gtrsim 1.5\,\mathrm{eV}^2$ and mixing amplitude $\sin^2(2\theta) \simeq 0.115$[48]. Taken together these results may be explained by models that incorporate one or more sterile neutrinos that mix with the active ones[49]-[56] including perhaps non-standard interactions[57]; although, there is some tension in the sterile neutrino interpretation of short-baseline anomalies[58]. A comprehensive review of short baseline oscillation experiments summarizes their interpretation in terms of one or more generations of sterile neutrinos[59].

Recently it has been pointed out that the presence of sterile neutrinos may induce a modification of the recently measured angle θ_{13} [119, 120].

Hints for the existence of sterile neutrinos also emerge from cosmology. The analysis of the cosmic microwave background anisotropies by WMAP[61] suggests that the effective number of neutrino species is $N_{eff} = 3.84 \pm 0.40$ and $\sum(m_\nu) < 0.44\,eV$, suggesting the case for sterile neutrino(s) with $m \lesssim$ eV, however the recent results from (SPT), (ACT)[62] and PLANCK[63] weaken the bounds considerably. Complementary cosmological data suggests that $N_{eff} > 3$ at the 95% confidence level[66], although accommodating an eV sterile neutrino requires a reassessment of other cosmological parameters[67]. For recent reviews on "light" sterile neutrinos see ref.[68]. Sterile neutrinos with masses in the $\sim$ keV range *may* also be suitable warm dark matter candidates[69]-[74] and appealing models of sterile neutrinos provide tantalizing mechanisms for baryogenesis[121].

These hints motivate several experimental proposals to search for sterile neutrinos (see the reviews in ref.[68]). Various experimental searches have been proposed, such as Higgs decay and matter interactions of relic sterile neutrinos[75], the end point of β-decay in ^{187}Re with a value of $Q = 2.5\,\mathrm{keV}$[76, 77], electron capture decays of $^{163}Ho \rightarrow^{163} Dy$[78] and ^{8}Li production and decay[122]. More recently, the focus has turned on the possible new facilities at the "intensity frontier" such as project X at Fermilab[79], alternative high intensity sources[68, 80] and recent proposals to study sterile-active oscillations with pion and kaon decay at rest (DAR)[81, 82] or muons from a storage ring[83] as well as the possibility of

discrimination between heavy Dirac and Majorana sterile neutrinos via $|\Delta L| = 2$ processes in high luminosity experiments[222], which is summarized in a recent review[59]. Although the recently reported analysis of the phase II data of the Mainz Neutrino Mass Experiment[123] found no evidence for a fourth neutrino state tightening the limits on the mass and mixing of a fourth sterile species, the possibility of a *heavy* sterile species is still actively explored[124, 125]. More recently the PIENU collaboration at TRIUMF[126] has reported an upper limit on the neutrino mixing matrix element $|U_{ei}|^2 \leq 10^{-8}$ $(90\%\, C.L.)$ in the neutrino mass region $60 - 129\,\mathrm{MeV}/c^2$.

In this article we focus on complementary consequences of sterile neutrinos in the form of *charged lepton mixing phenomena*. The discussion of whether or not charged leptons *oscillate* has been controversial[127]-[132], and more recently this question was addressed from the point of view of coherence[133] highlighting that while oscillations are possible, they lead to rapid decoherence and no observable effects. Muon-antimuon oscillations via massive Majorana neutrinos have been studied in ref.[134], however, to the best or our knowledge the issue of charged lepton $(\mu - e)$ *mixing* (we emphasize mixing over oscillations), has not yet received the same level of attention. Although in ref.[135] charged lepton mixing and oscillations as a consequence of neutrino mixing was studied in early Universe cosmology at temperatures $m_\mu \ll T \ll M_W$ where it was argued that medium effects enhance charged lepton mixing, the question of charged lepton *mixing* in vacuum and as a consequence of possible new generations of sterile neutrinos has not yet been discussed in the literature and is the main motivation of this article.

Furthermore we discuss the relationship between the lepton flavor violating decay $\mu \to e\gamma$, and charged lepton *mixing* in terms of self-energies and propagators that mix μ and e. Charged lepton violation is the focus of current experimental searches[136, 137], and a recent experimental proposal[138] to search for charged lepton flavor violation via the coherent conversion process $\mu - N \to e - N$ at Fermilab.

Goals: In this article we study both charged lepton *oscillations* and *mixing* as a consequence of intermediate states of mixed massive neutrinos, and discuss the relationship between charged lepton mixing and charged lepton flavor violating processes.

- **a) Oscillations:** In a recent article[139] (see also [85, 175]) we have provided a non-perturbative quantum field theoretical generalization of the Weisskopf-Wigner method to understand the correlated quantum state of charged leptons and neutrinos that consistently describes pion/kaon decay in real time. Knowledge of this state allows us to obtain the reduced density matrix for charged leptons by tracing out the neutrino degrees of freedom. The off diagonal density matrix elements in the flavor basis contains all the information on charged lepton (μ, e) *coherence and oscillations*.
- **b) Mixing:** Charged lepton oscillations evidenced in the reduced density matrix are a consequence of a common set of intermediate states that couple to the charged leptons. We then study the charged current contribution to the one-loop self-energy which couples charged leptons to an intermediate state of mixed massive neutrinos. The self-energy unambiguously determines the propagating states and explicitly describe charged lepton *mixing*. We obtain the mixed propagator, extracting the mixing angles and analyze the propagating modes and their wavefunctions. These results motivate us to address the relation between lepton flavor violating transitions such as $\mu \to e\gamma$ and charged lepton *mixing*.

Brief summary of results:

- **a:)** The quantum state of charged leptons and neutrinos from (light) pseudoscalar decay is a correlated entangled state from which we construct the corresponding (pure state) density matrix. Under the condition that neutrinos are not observed, we trace over their degrees of freedom leading to a reduced density matrix for the charged leptons. Because we focus solely on decay of π, K these are μ, e. Integrating out the unobserved neutrinos leads to a reduced density matrix that is off-diagonal in the flavor basis. The off-diagonal matrix elements describe charged lepton mixing and exhibit oscillations with typical frequency $E_\mu(k) - E_e(k) \gtrsim \mathcal{O}(m_\mu - m_e) \sim m_\mu \sim 1.6 \times 10^{23} s^{-1}$ which are unobservable over any experimentally relevant time scale and lead to rapid decoherence. This conclusion agrees with a similar observation in ref.[133]. While these fast oscillations lead to decoherence over microscopic time scales, we recognize that the *origin* of these oscillations are a common set of intermediate states akin to neutral meson oscillations.

- Recognizing that the origin of oscillations are intermediate states that are common to both charged leptons we obtain the self-energy contributions and the full mixed propagator for the μ, e system. Mixing is a direct result of charged current interactions with intermediate neutrino *mass eigenstates*. As in the case of neutral meson mixing we identify "short" and "long" distance contributions to the flavor off-diagonal self-energies. The "short" distance contribution corresponds to the intermediate state of a $W^{\pm}$ and neutrino mass eigenstates and is dominant, whereas the "long" distance contribution is described by an intermediate state of π, K and a neutrino mass eigenstate. We calculate explicitly the short distance and estimate the long distance contributions. Unitarity of the neutrino mixing matrix entails a Glashow-Ilioupoulos-Maiani (GIM) type mechanism that suppresses charged lepton mixing for light or nearly degenerate neutrinos, thus favoring heavy sterile neutrinos as intermediate states.

- We obtain the flavor off-diagonal charged lepton propagator and analyze in detail the propagating modes. $\mu - e$ mixing cannot be described solely in terms of a local off-diagonal mass matrix but also off-diagonal *kinetic terms* which are four-momentum dependent and contribute to off-shell processes. Mixing angles are GIM suppressed and both *chirality and four momentum dependent*. The largest angle corresponds to the negative chirality component, the difference in mixing angles near the muon and electron mass shells is independent of the local renormalization counterterms and is given by $\theta_L(M_\mu^2) - \theta_L(M_e^2) \propto G_F \sum_j U_{\mu j} U^*_{je} m_j^2$ where m_j is the mass of the intermediate neutrino. Therefore charged lepton mixing is dominated by intermediate states with mixed heavy neutrinos. Assuming one generation of a heavy sterile neutrino with mass M_S and extrapolating recent results from TRIUMF[126] we obtain an upper bound $\theta_L(M_\mu^2) - \theta_L(M_e^2) \lesssim 10^{-14} \left(M_S/100\,\mathrm{MeV} \right)^2$. We obtain the propagating eigenstates of charged leptons via two complementary methods: by direct diagonalization of the propagator and by field redefinitions followed by bi-unitary transformations, both results agree and yield momentum and chirality dependent mixing angles which are widely different on the respective mass shells.

- The relationship between charged lepton *mixing* and the lepton flavor violating decay $\mu \to e\gamma$ is discussed in terms of the mixed charged lepton self-energies and *possible*

observational effects in the form of further contributions to $\mu \to e\gamma$ are discussed. In particular we argue that writing the flavor lepton fields in terms of the propagating modes in flavor diagonal interaction vertices leads to novel interactions that depend on the difference of the mixing angles on the mass shells, this difference being independent of the choice of local renormalization counterterms.

3.2 REDUCED DENSITY MATRIX: CHARGED LEPTON OSCILLATIONS

In ref.[139] the quantum field theoretical Weisskopf-Wigner (non-perturbative) method has been implemented to obtain the quantum state resulting from the decay of a pseudoscalar meson M, (pion or kaon). It is found that such state is given by (see [139] for details and conventions),

$$|M^-_{\vec{p}}(t))\rangle = e^{-iE_M(p)t}e^{-\Gamma_M(p)\frac{t}{2}}|M^-_{\vec{p}}(0)\rangle - |\Psi_{l,\nu}(t)\rangle \tag{3.2.1}$$

where $|\Psi_{l,\nu}(t)\rangle$ is the *entangled state* of charged leptons and neutrinos given by

$$|\Psi_{l,\nu}(t)\rangle = \sum_{j,\alpha,\vec{q},h,h'} \left\{ U_{\alpha j}\, C_{\alpha j}(\vec{k},\vec{q},h,h';t)|l^-_\alpha(h,\vec{k})\rangle\, |\overline{\nu}_j(h',-\vec{q})\rangle \right\} \;\; ; \;\; \vec{k} = \vec{p} + \vec{q}\,, \tag{3.2.2}$$

where

$$C_{\alpha,j}(\vec{k},\vec{q},h,h';t) = \Pi_{\alpha j}\, \mathcal{M}_{\alpha j}(\vec{k},\vec{q},h,h')\, \mathcal{F}_{\alpha j}[\vec{k},\vec{q};t]\, e^{-i(E_\alpha(k)+E_j(q))t} \tag{3.2.3}$$

with

$$\mathcal{F}_{\alpha j}[\vec{q},\vec{p},h,h';t] = \left[\frac{1 - e^{-i(E^r_M(p)-E_\alpha(k)-E_j(q)-i\frac{\Gamma_M}{2})t}}{E^r_M(p) - E_\alpha(k) - E_j(q) - i\frac{\Gamma_M}{2}} \right] \tag{3.2.4}$$

and $\mathcal{M}_{\alpha,j}(\vec{k},\vec{q},h,h'), \Pi_{\alpha,j}(q,k)$ are the production matrix elements and phase space factors respectively,

$$\mathcal{M}_{\alpha,j}(\vec{k},\vec{q},h,h') = F_M\, \overline{\mathcal{U}}_{\alpha,h}(\vec{k})\gamma^\mu \mathbb{L} \mathcal{V}_{j,h'}(\vec{q}) p_\mu \tag{3.2.5}$$

$$\Pi_{\alpha,j}(q,k) = \frac{1}{\sqrt{8VE_M(p)E_\alpha(k)E_j(q)}} \tag{3.2.6}$$

In these expressions F_M is the pion or kaon decay constant, $\overline{\mathcal{U}}; \mathcal{V}_{j,h'}(\vec{q})$ are the spinors corresponding to the charged lepton α and the neutrino mass eigenstate j (for notation and details see ref.[139]). The *leptonic* density matrix that describes the *pure* quantum entangled state of neutrinos and charged leptons is given by

$$\rho_{l,\nu}(t) = |\Psi_{l,\nu}(t)\rangle\langle\Psi_{l,\nu}(t)| \tag{3.2.7}$$

If the neutrinos are *not observed* their degrees of freedom must be traced out in the density matrix, the resulting density matrix is no longer a pure state,

$$\rho_l^R(t) = \mathrm{Tr}_\nu \rho_{l,\nu}(t) = \sum_{j,\alpha,\cdots}\sum_{j',\beta,\cdots} U_{\alpha j}\, U^*_{j\beta} C_{\alpha,j} C^*_{\beta,j} |l^-_\alpha\rangle\langle l^-_\beta| \langle\overline{\nu}_j|\overline{\nu}_{j'}\rangle\,, \tag{3.2.8}$$

where $\langle\overline{\nu}_j|\overline{\nu}_{j'}\rangle = \delta_{jj'}$.

Considering only light pseudoscalar decay π, K, the only charged leptons available are μ, e. For a fixed helicity h and momentum $\vec{k}$ of the charged leptons the reduced density matrix is given by

$$\begin{aligned} \rho_l^R(t) &= \rho_{ee}(h,\vec{k},t)|e^-_{h,\vec{k}}\rangle\ \langle e^-_{h,\vec{k}}| + \rho_{\mu\mu}(h,\vec{k},t)|\mu^-_{h,\vec{k}}\rangle\ \langle\mu^-_{h,\vec{k}}| \\ &+ \rho_{e\mu}(h,\vec{k},t)|e^-_{h,\vec{k}}\rangle\ \langle\mu^-_{h,\vec{k}}| + \rho_{\mu,e}(h,\vec{k},t)|\mu^-_{h,\vec{k}}\rangle\ \langle e^-_{h,\vec{k}}| \end{aligned} \tag{3.2.9}$$

The diagonal density matrix elements in the μ, e basis describe the *population* of the produced charged leptons whereas the off-diagonal elements describe the *coherences*. The diagonal matrix elements $\rho_{\alpha\alpha}$, $\alpha = \mu, e$ are given by

$$\rho^R_{\alpha\alpha}(t) = \sum_j |U_{\alpha,j}|^2\, \mathrm{BR}_{M\to l_\alpha \nu_j}\left[1 - e^{-\Gamma_M t}\right] \ \ ; \ \ \alpha = \mu, e \tag{3.2.10}$$

where BR are the branching ratios $\Gamma_{M\to l_\alpha\ \nu_j}/\Gamma_M$ and we have used some results obtained in ref.[139]. The off diagonal elements do not have a simple expression, however the most important aspect for the discussion is that these density matrix elements are of the form

$$\rho^R_{\mu e} = \sum_j U_{\mu j}\, U^*_{je} C_{\mu,j} C^*_{e,j}\ ;\ \rho^R_{e\mu} = (\rho^R_{\mu e})^*\,, \tag{3.2.11}$$

where the coefficients $C_{\alpha,j}$ are given by (3.2.3-3.2.6). These matrix elements describe the process $M \to \alpha\nu_j$ followed by a "recombination"-type process $M\nu_j \to \beta$ thereby suggesting the intermediate state $M\nu_j$ common to both matrix elements. For $\Gamma_M t \gg 1$ the reduced density matrix (for fixed h, k) in the charged lepton basis is of the form

$$\rho^R = \begin{bmatrix} A_{ee} & A_{\mu e}e^{-i(E_\mu(k)-E_e(k))t} \\ A_{e\mu}e^{i(E_\mu(k)-E_e(k))t} & A_{\mu\mu} \end{bmatrix}, \tag{3.2.12}$$

This tells us that there will be $\mu \Leftrightarrow e$ oscillations. However, these oscillations occur with large frequencies $E_\mu(k) - E_e(k) \gtrsim \mathcal{O}(m_\mu - m_e) \sim m_\mu \sim 1.6 \times 10^{23} s^{-1}$ and are unobservable over any experimentally relevant time scale. This conclusion agrees with a similar observation in ref.[133].

Although these oscillations average out over relevant time scales and are experimentally unobservable, an important issue is *their origin*. The mixing between charged leptons arises from the fact that they share common *intermediate states*, in the case studied above the common intermediate state corresponds to a pseudoscalar meson and a *neutrino mass eigenstate*.

Two aspects are important in the off diagonal terms in (3.2.11) whose long time limit defines $A_{\mu e}$: a) from the expression (3.2.11) it follows that $A_{\mu e} \propto F_M^2$ and b) if all the neutrino states are degenerate the off diagonal terms vanish because the $C_{\alpha,j}$ would be the same for all j and $\sum_j U_{\mu j}U_{je}^* = 0$ by unitarity of the mixing matrix. This cancellation for massless or degenerate neutrinos is akin to the GIM mechanism.

From this point of view the physical origin of the oscillations is found in *mixing* of the charged leptons from the fact that they share common intermediate states. This is in fact similar to the oscillations and mixing through radiative corrections with common intermediate states in the $K_0\overline{K}_0$ system. The obvious difference with this system is that, in absence of weak interactions, K_0 and $\overline{K}_0$ are degenerate and this degeneracy is lifted by the coupling to the (common) intermediate states, leading to oscillations on long time scales.

The conclusion of this discussion is that charged lepton oscillations are a result of their *mixing* via a set of common intermediate states. The off-diagonal density matrix elements

are of $\mathcal{O}(F_M^2)$, these are the lowest order corrections in a perturbative expansion, therefore they do not reveal the full structure of the mixing phenomenon.

If charged leptons mix via a common set of intermediate states, the correct propagating degrees of freedom are described by poles in the full charged lepton propagator which requires the self-energy correction. Such self-energy will reflect the *mixing* through the intermediate states. Whereas oscillations average out the off-diagonal density matrix on short time scales, the main physical phenomenon of mixing is manifest in the true propagating modes, namely the poles in the propagator which now becomes an off diagonal matrix in flavor space.

3.3 CHARGED LEPTON MIXING:

We argued above that lepton mixing is a consequence of an intermediate meson/neutrino state which couples to *both* charged leptons. The intermediate meson state is a low energy or "long distance" representation of the coupling of charged leptons to quarks via charged current interactions and is akin to the mixing between $K_0\overline{K}_0$ via intermediate states with two and three pions. This "long distance" (low energy) contribution to the charged lepton self energy is depicted in fig. (17).

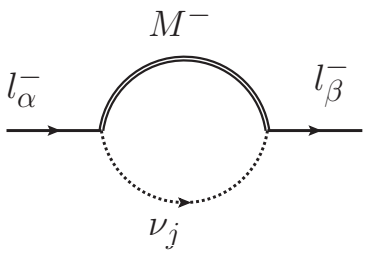


Figure 17: Long distance contribution: intermediate state with ν_j and $M = \pi, K$.

This is a low energy representation of physical process in which a lepton couples to an intermediate W vector boson and a neutrino mass eigenstate, followed by the decay of the (off-shell) W into quark-antiquark pairs with the quantum numbers of the pseudoscalar

mesons. Therefore we also expect a *short distance* contribution in which the intermediate state corresponds simply to the exchange of a W boson and a neutrino mass eigenstate. This contribution to the charged lepton self-energy is depicted in fig. (18).

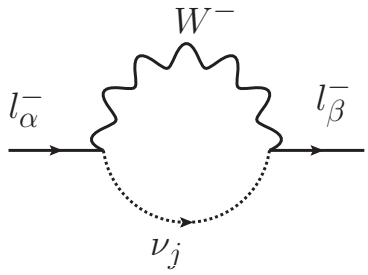


Figure 18: Short distance contribution: intermediate state with ν_j and W^-.

We shall calculate both self-energy diagrams to properly ascertain each contribution to the full charged lepton propagator.

However, before carrying out the detailed calculations we point out that there are also electromagnetic and neutral current contributions to the self-energies. However these are flavor *diagonal*, thus while they will both contribute to the self-energies, only the charged current contributions (long and short distance) lead to *off diagonal* self energies which lead to charged lepton *mixing*. Furthermore, both long and short distance self-energies are of the general form

$$\Sigma_{\alpha\beta} \propto \sum_j U_{\alpha j}\, S_j\, U^*_{j\beta} \tag{3.3.1}$$

where S_j is the propagator of neutrino mass eigenstates, therefore unitarity of the mixing matrix $\sum_j U_{\alpha j} U^*_{j\beta} = \delta_{\alpha\beta}$ leads to a GIM (Glashow-Ilioupoulos-Maiani) type-suppression of the off-diagonal matrix elements: if all the neutrinos in the loop are degenerate, unitarity entails that there is no off-diagonal contribution to the self-energy, furthermore, this argument also suggests that the off-diagonal terms will be dominated by the most massive neutrino state.

Therefore charged lepton *mixing* is a consequence of off-diagonal components of the self-energy matrix, which results in an off-diagonal *propagator* for $e-\mu$ leptons as a consequence of common neutrino mass eigenstates in intermediate states. If the neutrinos are either

massless or degenerate the unitarity of the mixing matrix leads to vanishing off-diagonal matrix elements and no mixing.

3.3.1 Short distance contribution:

We begin by computing the self energy contribution from W exchange depicted in fig. (18). Throughout this calculation, we shall be working in the physical unitary gauge and in dimensional regularization. Upon passing to the basis of mass eigenstates that define the neutrino propagators, $\psi_{\nu_\alpha} = \sum_j U_{\alpha j}\psi_j$ the charged current contribution to the self energy *matrix* is given by

$$-i\Sigma_{\alpha\beta} = \left(\frac{-ig}{\sqrt{2}}\right)^2 \sum_j \int \frac{d^4k}{(2\pi)^4} U_{\alpha j}\gamma_\mu \mathbb{L}\left(\frac{i(\not{k}+m_j)}{k^2-m_j^2+i\epsilon}\right) U^*_{\beta j}\gamma_\nu \mathbb{L}\left[\frac{-i\left(g^{\mu\nu}-\frac{k^\mu k^\nu}{M_W^2}\right)}{(p-k)^2-M_W^2}\right] \quad (3.3.2)$$

where $\mathbb{L},\mathbb{R} = (1\mp\gamma^5)/2$ respectively.

This integral is calculated in dimensional regularization. We introduce a renormalization scale κ which we choose $\kappa = M_W$ thus renormalizing at the $W-pole$, and define

$$\Delta_j = -p^2x(1-x) + m_j^2 x + M_W^2(1-x)\,, \quad (3.3.3)$$

separating explicitly the divergent and finite parts in the $\overline{\text{MS}}$ scheme we find

$$\Sigma_{\alpha\beta}(p) = \not{p}\,\mathbb{L}\left[\sum_j U_{\alpha j}U^*_{j\beta}\int_0^1 \left[I_j^d(p^2;x) + I_j^f(p^2;x)\right]\,dx\right] \quad (3.3.4)$$

where

$$I_j^d(p^2;x) = -\frac{g^2\,M_W^{-\epsilon}}{2(4\pi)^2}(1-x)\left[2+\frac{3\Delta_j}{M_W^2}+(1-x)^2\frac{p^2}{M_W^2}\right]\left(\frac{2}{\epsilon}-\gamma+\ln 4\pi\right) \quad (3.3.5)$$

$$I_j^f(p^2;x) = \frac{g^2 M_W^{-\epsilon}}{2(4\pi)^2}(1-x)\left[\left(2+\frac{3\Delta_j}{M_W^2}+(1-x)^2\frac{p^2}{M_W^2}\right)\ln\frac{\Delta_j}{M_W^2}+\frac{2xm_j^2}{M_W^2}\right]. \quad (3.3.6)$$

Unitarity of the neutrino mixing matrix in the form $\sum_j U_{\alpha j} U^*_{j\beta} = \delta_{\alpha\beta}$ leads to GIM-like cancellations in both the divergent and the finite parts for the off-diagonal components of the self-energy matrix. Therefore for $\alpha \neq \beta$ we find

$$\Sigma_{\alpha\beta}(p) = \not{p}\,\mathbb{L}\left[z^d_{\alpha\beta} + z^f_{\alpha\beta}(p^2)\right], \tag{3.3.7}$$

where

$$z^d_{\alpha\beta} = -\frac{g^2\, M_W^{-\epsilon}}{64\pi^2} \sum_j U_{\alpha j} U^*_{j\beta} \frac{m_j^2}{M_W^2} \left(\frac{2}{\epsilon} - \gamma + \ln 4\pi\right) \quad ; \quad \alpha \neq \beta \tag{3.3.8}$$

and

$$z^f_{\alpha\beta}(p^2) = \frac{g^2\, M_W^{-\epsilon}}{32\pi^2} \quad \sum_j U_{\alpha j} U^*_{j\beta} \int_0^1 \left\{ \left[2 + \frac{3\Delta_j}{M_W^2} + (1-x)^2 \frac{p^2}{M_W^2}\right] \ln\left(\frac{\Delta_j}{M_W^2}\right) + \frac{2m_j^2}{M_W^2} x \right\} (1-x)\, dx \quad ; \quad \alpha \neq \beta\,. \tag{3.3.9}$$

3.3.2 Long distance contribution:

We now turn our attention to the intermediate state described by the exchange of a π/K meson and a neutrino mass eigenstate. This is the state that suggested charged lepton mixing from the density matrix treatment in the previous section. A difficulty arises in the calculation of the meson exchange because in order to properly describe the coupling between the meson and the charged lepton and neutrinos we would need the full off-shell form factor $F_M(q^2)$ which is a function of the loop momentum since the meson is propagating off its mass shell in the intermediate state. Clearly this is very difficult to include in a reliable calculation, therefore we restrict our study to an *estimate* of this contribution obtained by simply using the on-shell value of the form factor, namely the meson decay constant F_M in order to obtain an admittedly rough assessment of its order of magnitude.

Under this approximation the contribution to the self-energy matrix from this intermediate state is given by

$$-i\Sigma^M_{\alpha\beta} = F_M^2 \sum_j U^*_{\alpha j} U_{\beta j} \int \frac{d^4k}{(2\pi)^4} (\not{p} - \not{k}) \mathbb{L} \left(\frac{i(\not{k} + m_j)}{k^2 - m_j^2 + i\epsilon}\right) (\not{p} - \not{k}) \mathbb{L} \left(\frac{i}{(p-k)^2 - M_M^2 + i\epsilon}\right) \tag{3.3.10}$$

where M_M is the meson mass. The width of the meson may be incorporated via a Breit-Wigner approximation $M_M \to M_M - i\Gamma_M/2$, however this will only yield a contribution which is higher order in G_F.

The calculation is performed in dimensional regularization, choosing the renormalization scale $\kappa = M_W$ as for the short distance contribution, introducing

$$\delta_j = -p^2 x(1-x) + m_j^2 x + M_M^2(1-x)\,, \tag{3.3.11}$$

and separating the divergent and finite parts in the $\overline{MS}$ scheme we find

$$\Sigma_{\alpha\beta}(p) = \not{p}\,\mathbb{L}\left[\sum_j U_{\alpha j}U^*_{j\beta}\int_0^1 \Big[J_j^d(p^2;x) + J_j^f(p^2;x)\Big]\,dx\right], \tag{3.3.12}$$

where

$$J_j^d(p^2;x) = -\frac{M_W^{2-\epsilon}F_M^2}{(4\pi)^2}\Big(\delta_j(1+3x) - (1-x)x^2p^2\Big)\left(\frac{2}{\epsilon} - \gamma + \ln 4\pi\right), \tag{3.3.13}$$

$$J_j^f(p^2;x) = \frac{M_W^{2-\epsilon}F_M^2}{(4\pi)^2}\left[2x^2\frac{\delta_j}{M_W^2} + \left((1-x)x^2\frac{p^2}{M_W^2} - (1+3x)\,\frac{\delta_j}{M_W^2}\right)\,\ln\frac{\delta_j}{M_W^2}\right]. \tag{3.3.14}$$

For the off-diagonal matrix elements, unitarity of the neutrino mixing matrix leads to GIM type cancellations as in the short distance case, therefore for $\alpha \neq \beta$ we find

$$\Sigma_{\alpha\beta}(p) = \not{p}\,\mathbb{L}\left[\varsigma^d_{\alpha\beta} + \varsigma^f_{\alpha\beta}(p^2)\right], \tag{3.3.15}$$

where

$$\varsigma^d_{\alpha\beta} = -3\frac{M_W^{2-\epsilon}F_M^2}{32\pi^2}\,\sum_j U_{\alpha j}U^*_{j\beta}\frac{m_j^2}{M_W^2}\,\left(\frac{2}{\epsilon} - \gamma + \ln 4\pi\right) \quad ; \quad \alpha \neq \beta \tag{3.3.16}$$

$$\begin{aligned}\varsigma^f_{\alpha\beta}(p^2) \;&=\; \frac{M_W^{2-\epsilon}F_M^2}{16\pi^2}\,\sum_j U_{\alpha j}U^*_{j\beta}\int_0^1\left[2x^2\frac{\delta_j}{M_W^2} + \Big((1-x)x^2\frac{p^2}{M_W^2} - (1+3x)\,\frac{\delta_j}{M_W^2}\Big)\,\ln\frac{\delta_j}{M_W^2}\right]dx \\ &\alpha \neq \beta\end{aligned} \tag{3.3.17}$$

However, with $F_M \propto G_F f_{\pi,K}$ and $f_{\pi,K} \sim 100\,\mathrm{MeV}$ it follows that

$$F_M^2\,M_W^2 \propto g^2\Big(\frac{g f_{\pi,k}}{M_W}\Big)^2 \sim 10^{-8}g^2 \tag{3.3.18}$$

therefore the long distance contribution is negligible as compared to the short distance contribution and to leading order the off-diagonal components of the self energy are given by eqns. (3.3.7- 3.3.9).

As noted previously unitarity of the neutrino mixing matrix entails that the flavor off diagonal matrix elements of the self-energy vanish either for vanishing or degenerate neutrino masses. Obviously the contribution from light active-like neutrinos is strongly suppressed by the ratios m_j^2/M_W^2, hence these off-diagonal matrix elements are dominated by the *heaviest* species of sterile neutrinos.

Thus charged lepton mixing is enhanced by intermediate states with heavy sterile neutrinos. This is one of the main results of this article.

If even the heaviest generation of sterile neutrinos feature masses $m_j \ll M_W$ and for $p^2 \ll M_W^2$ the following order of magnitude for the off-diagonal component $z_{\mu e}$ is obtained

$$z_{\mu e} \simeq \frac{G_F}{4\pi^2} \sum_j U_{\alpha j} U^*_{j\beta}\, m_j^2 \,, \tag{3.3.19}$$

as it will be seen below this estimate determines the mixing angles up to kinematic factors.

3.4 FULL PROPAGATOR: MIXING ANGLES AND PROPAGATING MODES.

3.4.1 Full propagator and mixing angles:

To treat μ, e mixing it is convenient to introduce a flavor doublet

$$\Psi = \begin{pmatrix} \psi_\mu \\ \psi_e \end{pmatrix} , \tag{3.4.1}$$

The general structure of the self-energy is of the form

$$\mathbf{\Sigma}(p) = \Big[\mathbf{z}_L(p^2)\not{p} + \delta\mathbf{M}_L(p^2)\Big]\mathbb{L} + \Big[\mathbf{z}_R(p^2)\not{p} + \delta\mathbf{M}_R(p^2)\Big]\mathbb{R} \,. \tag{3.4.2}$$

The neutral current interactions contributes generally to the right and left components of the self-energy but are *diagonal* in flavor and so are the electromagnetic contributions. The $V-A$ nature of the charged current interactions is such that their contribution is only of the form $\mathbf{z}_L(p^2)\not{p}\,\mathbb{L}$ and is the *only* contribution that yields flavor off-diagonal terms and are ultimately responsible for $\mu - e$ mixing. To cancel the poles in ϵ in the self-energy we allow counterterms in the bare Lagrangian

$$\mathcal{L}_{ct} = \overline{\Psi}(\delta\mathbf{Z}_{ct} - 1)\not{p}\Psi + \overline{\Psi}\delta\mathbf{M}\Psi + \text{h.c.}\,. \tag{3.4.3}$$

The full propagator $\mathbf{S}$ now becomes a 2×2 matrix which is the solution of

$$\big[\not{p}\mathbf{1} + \not{p}(\delta\mathbf{Z}_{ct} - 1) - \mathbf{\Sigma}(p) - \mathbf{M}\big]\mathbf{S} = \mathbf{1} \tag{3.4.4}$$

where the boldfaced quantities are 2×2 matrices and

$$\mathbf{M} = \begin{pmatrix} M_\mu & 0 \\ 0 & M_e \end{pmatrix}. \tag{3.4.5}$$

In what follows we will assume that that $\mathbf{M}$ contains the renormalized masses and we will neglect finite momentum dependent contributions to $\mathbf{M}$ since these will only generate higher order corrections to the mixing matrix as will become clear below.

We will choose the counterterm $(\delta\mathbf{Z}_{ct} - 1)$ in the $\overline{MS}$ scheme to cancel the term $z^d_{\alpha\beta}$ in eqn. (3.3.7). Therefore equation (3.4.4) becomes

$$\Big[\not{p}\,\mathbf{Z}_L^{-1}\,\mathbb{L} + \not{p}\,\mathbf{Z}_R^{-1}\,\mathbb{R} - \mathbf{M}\Big]\mathbf{S} = \mathbf{1} \tag{3.4.6}$$

where

$$\mathbf{Z}^{-1}_{L,R} = \mathbf{1} - \mathbf{z}^f_{L,R}(p^2)\,. \tag{3.4.7}$$

The leading contribution to the *off-diagonal* matrix elements is given by the "short-distance" term eqn. (3.3.9).

Multiplying on the left both sides of (3.4.6) by $\not{p} + \mathbf{M}\mathbf{Z}_R\,\mathbb{L} + \mathbf{M}\mathbf{Z}_L\,\mathbb{R}$ and writing the full propagator as

$$\mathbf{S} = \mathbb{R}\mathbf{S}_R + \mathbb{L}\,\mathbf{S}_L \tag{3.4.8}$$

where

$$\mathbf{S}_R = \mathbf{A}_R(p^2)\Big[\not{p} + \mathbf{B}_R(p^2)\Big] \tag{3.4.9}$$

$$\mathbf{S}_L = \mathbf{A}_L(p^2)\Big[\not{p} + \mathbf{B}_L(p^2)\Big] \tag{3.4.10}$$

we find

$$\Big(p^2\,\mathbf{Z}_R^{-1} - \mathbf{M}\mathbf{Z}_L\mathbf{M}\Big)\mathbf{A}_R(p^2) = \mathbf{1} \tag{3.4.11}$$

$$\Big(p^2\,\mathbf{Z}_L^{-1} - \mathbf{M}\mathbf{Z}_R\mathbf{M}\Big)\mathbf{A}_L(p^2) = \mathbf{1} \tag{3.4.12}$$

and the conditions

$$\mathbf{B}_R(p^2) = \mathbf{M}\,\mathbf{Z}_L(p^2) \quad ; \quad \mathbf{B}_L(p^2) = \mathbf{M}\,\mathbf{Z}_R(p^2)\,. \tag{3.4.13}$$

In what follows we will neglect CP violating phases in $U_{\alpha j}$ with the purpose of studying $\mu - e$ mixing in the simplest case. Under these approximations we find

I): The solution for $\mathbf{A}_R(p^2)$ in eqn. (3.4.11) is obtained as follows. Consider the diagonalization of the inverse propagator

$$p^2\,\mathbf{Z}_R^{-1} - \mathbf{M}\mathbf{Z}_L\mathbf{M} = \frac{1}{2}\Big[Q_\mu^R(p^2) + Q_e^R(p^2)\Big]\mathbb{1} - \frac{\lambda_R(p^2)}{2}\begin{pmatrix} \cos 2\,\theta_R(p^2) & \sin 2\,\theta_R(p^2) \\ \sin 2\,\theta_R(p^2) & -\cos 2\,\theta_R(p^2) \end{pmatrix} \tag{3.4.14}$$

where

$$Q_\alpha^R(p^2) = p^2\big[\mathbf{Z}_R^{-1}\big]_{\alpha\alpha} - M_\mu^2\big[\mathbf{Z}_L\big]_{\alpha\alpha} \quad ; \quad \alpha = \mu\,;\, e \tag{3.4.15}$$

and

$$\lambda_R(p^2) = \left[\Big(Q_\mu^R(p^2) - Q_e^R(p^2)\Big)^2 + 4\Big(M_\mu M_e\, z^f_{L,\mu e}(p^2)\Big)^2\right]^{\frac{1}{2}}. \tag{3.4.16}$$

To leading order we find the mixing angle to be given by

$$\tan 2\,\theta_R(p^2) = \frac{2M_\mu M_e\, z^f_{L,\mu e}(p^2)}{M_\mu^2 - M_e^2}\,. \tag{3.4.17}$$

The matrix above can be diagonalized by a unitary transformation

$$\mathcal{U}[\theta] = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{3.4.18}$$

in terms of the mixing angle $\theta_R(p^2)$, namely

$$\mathcal{U}[\theta_R(p^2)]\Big[p^2\,\mathbf{Z}_R^{-1} - \mathbf{M}\mathbf{Z}_L\mathbf{M}\Big]\mathcal{U}^{-1}[\theta_R(p^2)] = \begin{pmatrix} Q_\mu^R(p^2) - \varrho_R(p^2) & 0 \\ 0 & Q_e^R(p^2) + \varrho_R(p^2) \end{pmatrix} \tag{3.4.19}$$

where to leading order

$$\varrho_R(p^2) = \frac{1}{2}(M_\mu^2 - M_e^2)\tan^2 2\theta_R(p^2)\,, \tag{3.4.20}$$

leading to the result

$$\mathbf{A}_R(p^2) = \mathcal{U}^{-1}[\theta_R(p^2)] \begin{pmatrix} \frac{1}{Q_\mu^R(p^2)-\varrho_R(p^2)+i\epsilon} & 0 \\ 0 & \frac{1}{Q_e^R(p^2)+\varrho_R(p^2)+i\epsilon} \end{pmatrix} \mathcal{U}[\theta_R(p^2)]\,, \tag{3.4.21}$$

which (to leading order) simplifies to

$$\mathbf{A}_R(p^2) \simeq \mathcal{U}^{-1}[\theta_R(p^2)] \begin{pmatrix} \frac{Z_{\mu\mu}^R(p^2)}{p^2-M_\mu^2(p^2)-\varrho_R(p^2)+i\epsilon} & 0 \\ 0 & \frac{Z_{ee}^R(p^2)}{p^2-M_e^2(p^2)+\varrho_R(p^2)+i\epsilon} \end{pmatrix} \mathcal{U}[\theta_R(p^2)]\,. \tag{3.4.22}$$

In the above expressions $M_\mu^2(p^2), M_e^2(p^2)$ include the finite renormalization from the diagonal contributions of the self-energy matrix which have not been calculated here, furthermore the residues at the poles (wave-function renormalization) are also *finite* since the (local) divergent contributions are canceled by the counterterm.

Therefore $\mathbf{S}_R$ can be written in the basis that diagonalizes the kinetic term

$$\mathcal{U}[\theta_R(p^2)]\,\mathbf{S}_R\mathcal{U}^{-1}[\theta_R(p^2)] = \begin{pmatrix} \frac{Z_{\mu\mu}^R(p^2)\left[\not{p}+b_{\mu\mu}^R(p^2)\right]}{p^2-M_\mu^2(p^2)-\varrho_R(p^2)+i\epsilon} & \frac{Z_{\mu\mu}^R(p^2)\,b_{\mu e}^R(p^2)}{p^2-M_\mu^2(p^2)-\varrho_R(p^2)+i\epsilon} \\ \frac{Z_{ee}^R(p^2)\,b_{e\mu}^R(p^2)}{p^2-M_e^2(p^2)+\varrho_R(p^2)+i\epsilon} & \frac{Z_{ee}^R(p^2)\left[\not{p}+b_{ee}^R(p^2)\right]}{p^2-M_e^2(p^2)+\varrho_R(p^2)+i\epsilon} \end{pmatrix}\,, \tag{3.4.23}$$

where

$$\mathbf{b}^R(p^2) = \mathcal{U}[\theta_R(p^2)]\,\mathbf{M}\mathbf{Z}_L(p^2)\,\mathcal{U}^{-1}[\theta_R(p^2)]\,. \tag{3.4.24}$$

II): We proceed in the same manner for $\mathbf{A}_L(p^2)$, namely consider diagonalizing the inverse propagator

$$p^2\,\mathbf{Z}_L^{-1} - \mathbf{M}\mathbf{Z}_R\mathbf{M} = \frac{1}{2}\Big[Q_\mu^L(p^2) + Q_e^L(p^2)\Big]\mathbb{1} - \frac{\lambda_L(p^2)}{2}\begin{pmatrix} \cos 2\,\theta_L(p^2) & \sin 2\,\theta_L(p^2) \\ \sin 2\,\theta_L(p^2) & -\cos 2\,\theta_L(p^2) \end{pmatrix} \tag{3.4.25}$$

where

$$Q_\mu^L(p^2) = p^2\big[\mathbf{Z}_L^{-1}\big]_{\alpha\alpha} - M_\mu^2\big[\mathbf{Z}_R\big]_{\alpha\alpha} \quad ; \quad \alpha = \mu\,;\, e \tag{3.4.26}$$

and

$$\lambda_L(p^2) = \left[\Big(Q_\mu^L(p^2) - Q_e^L(p^2)\Big)^2 + 4\Big(p^2\,z_{L,\mu e}^f(p^2)\Big)^2\right]^{\frac{1}{2}}. \tag{3.4.27}$$

Again, to leading order we find the mixing angle to be given by

$$\tan 2\,\theta_L(p^2) = \frac{2p^2\,z_{L,\mu e}^f(p^2)}{M_\mu^2 - M_e^2}\,. \tag{3.4.28}$$

The matrix above can be diagonalized by the unitary transformation (3.4.18) now in terms of the mixing angle $\theta_L(p^2)$, namely

$$\mathcal{U}[\theta_L(p^2)]\Big[p^2\,\mathbf{Z}_L^{-1} - \mathbf{M}\mathbf{Z}_R\mathbf{M}\Big]\mathcal{U}^{-1}[\theta_L(p^2)] = \begin{pmatrix} Q_\mu^L(p^2) - \varrho_L(p^2) & 0 \\ 0 & Q_e^L(p^2) + \varrho_L(p^2) \end{pmatrix}, \tag{3.4.29}$$

where to leading

$$\varrho_L(p^2) = \frac{1}{2}(M_\mu^2 - M_e^2)\tan^2 2\theta_L(p^2) \tag{3.4.30}$$

leading to the result

$$\mathbf{A}_L(p^2) = \mathcal{U}^{-1}[\theta_L(p^2)]\begin{pmatrix} \frac{1}{Q_\mu^L(p^2) - \varrho_L(p^2) + i\epsilon} & 0 \\ 0 & \frac{1}{Q_e^L(p^2) + \varrho_L(p^2) + i\epsilon} \end{pmatrix}\mathcal{U}[\theta_L(p^2)]\,, \tag{3.4.31}$$

Neglecting the diagonal contributions to mass renormalization, but keeping the (finite) wave function renormalizations, the result (3.4.31) simplifies to

$$\mathbf{A}_L(p^2) \simeq \mathcal{U}^{-1}[\theta_L(p^2)]\begin{pmatrix} \frac{Z_{\mu\mu}^L(p^2)}{p^2 - M_\mu^2(p^2) - \varrho_L(p^2) + i\epsilon} & 0 \\ 0 & \frac{Z_{ee}^L(p^2)}{p^2 - M_e^2(p^2) + \varrho_L(p^2) + i\epsilon} \end{pmatrix}\mathcal{U}[\theta_L(p^2)]\,, \tag{3.4.32}$$

Just as in the previous case, $M^2_\mu(p^2), M^2_e(p^2)$ include the finite contribution from mass terms in the self energy and the residues at the poles are also finite, the local, divergent contribution being canceled by the counterterm.

The component $\mathbf{S}_L$ can now be written as

$$\mathcal{U}[\theta_L(p^2)]\,\mathbf{S}_L\,\mathcal{U}^{-1}[\theta_L(p^2)] = \begin{pmatrix} \frac{Z^L_{\mu\mu}(p^2)\left[\not{p}+b^L_{\mu\mu}(p^2)\right]}{p^2-M^2_\mu(p^2)-\varrho_L(p^2)+i\epsilon} & \frac{Z^L_{\mu\mu}(p^2)\,b^L_{\mu e}(p^2)}{p^2-M^2_\mu(p^2)-\varrho_L(p^2)+i\epsilon} \\ \frac{Z^L_{ee}(p^2)\,b^R_{e\mu}(p^2)}{p^2-M^2_e(p^2)+\varrho_L(p^2)+i\epsilon} & \frac{Z^L_{ee}(p^2)\left[\not{p}+b^L_{ee}(p^2)\right]}{p^2-M^2_e(p^2)+\varrho_L(p^2)+i\epsilon} \end{pmatrix}\,, \tag{3.4.33}$$

where

$$\mathbf{b}^L(p^2) = \mathcal{U}[\theta_L(p^2)]\,\mathbf{M}\mathbf{Z}_R(p^2)\,\mathcal{U}^{-1}[\theta_L(p^2)]\,. \tag{3.4.34}$$

An important aspect is that the mixing angles $\theta_R(p^2), \theta_L(p^2)$ not only are different for the R, L components a consequence of the $V-A$ nature of charged currents, but also that they feature very different momentum dependence,

$$\theta_R(p^2) \simeq \frac{M_e}{M_\mu} z^f_{L,\mu e}(p^2) \quad ; \quad \theta_L(p^2) \simeq \frac{p^2}{M^2_\mu} z^f_{L,\mu e}(p^2) \tag{3.4.35}$$

Near the muon mass shell $p^2 \simeq M^2_\mu$ it follows that $\theta_L \gg \theta_R$, for near the electron mass shell $p^2 \simeq M^2_e$ it follows that $\theta_R \gg \theta_L$. Off-shell, for virtuality $p^2 \gg M^2_\mu$ mixing of the L component becomes dominant.

In general the transformations that diagonalize the kinetic terms $\not{p}$ for both the positive and negative chirality components *do not* diagonalize the mass terms. In the basis in which the kinetic terms are diagonal the pole-structure of the propagator is revealed and the propagating modes can be read-off. This basis, however, does not diagonalize the mass term of the propagator and attempting to diagonalize the latter either via a unitary or a bi-unitary transformation will lead to an off diagonal matrix multiplying the kinetic term. A similar situation has been found in different contexts[142, 140, 141].

3.4.2 Propagating modes: the effective Dirac equation:

The nature of the propagating modes is best illuminated by solving the effective Dirac equation for the flavor doublet, which corresponds to the zeroes of the inverse propagator, namely

$$\Big[\not{p}\,\mathbf{Z}_L^{-1}\,\mathbb{L} + \not{p}\,\mathbf{Z}_R^{-1}\,\mathbb{R} - \mathbf{M}\Big]\,\Psi(p) = 0\,, \tag{3.4.36}$$

with Ψ a spinor doublet,

$$\Psi = \begin{pmatrix} \xi^R \\ \xi^L \end{pmatrix} \quad ; \quad \xi^{R,L} = \begin{pmatrix} \xi_\mu^{R,L} \\ \xi_e^{R,L} \end{pmatrix}\,. \tag{3.4.37}$$

It is convenient to work in the chiral representation and expand the positive and negative chirality components in the helicity basis

$$\vec{\sigma}\cdot\,\frac{\vec{p}}{|\vec{p}|}v_h(\vec{p}) = h v_h(\vec{p}) \quad ; \quad h = \pm 1 \tag{3.4.38}$$

in terms of which the spinor flavor doublet

$$\Psi(p) = \sum_h v_h \otimes \begin{pmatrix} \xi_h^R \\ \xi_h^L \end{pmatrix}\,, \tag{3.4.39}$$

where $\xi_h^{R,L}$ are flavor doublets that obey the following equations

$$\mathbf{Z}_L^{-1}(p_0 + p\,h)\xi_h^L + \mathbf{M}\xi_h^R = 0 \tag{3.4.40}$$

$$\mathbf{Z}_R^{-1}(p_0 - p\,h)\xi_h^R + \mathbf{M}\xi_h^L = 0\,. \tag{3.4.41}$$

The positive and negative energy and helicity components are given by ($p \equiv |\vec{p}|$)

$$\begin{pmatrix} \xi^R \\ -\frac{\mathbf{Z}_L\,\mathbf{M}}{p_0+p}\,\xi^R \end{pmatrix} \quad ; \quad p_0 > 0; h = 1 \quad ; \quad \begin{pmatrix} -\frac{\mathbf{Z}_R\,\mathbf{M}}{p_0+p}\,\xi^L \\ \xi^L \end{pmatrix} \quad ; \quad p_0 > 0, h = -1 \tag{3.4.42}$$

$$\begin{pmatrix} \frac{\mathbf{Z}_R\,\mathbf{M}}{|p_0|+p}\,\xi^L \\ \xi^L \end{pmatrix} \quad ; \quad p_0 < 0; h = 1 \quad ; \quad \begin{pmatrix} \xi^R \\ \frac{\mathbf{Z}_L\,\mathbf{M}}{|p_0|+p}\,\xi^R \end{pmatrix} \quad ; \quad p_0 < 0, h = -1 \tag{3.4.43}$$

The flavor doublets obey

$$\Big(p^2\,\mathbf{Z}_R^{-1} - \mathbf{M}\mathbf{Z}_L\mathbf{M}\Big)\xi^R(p) \;=\; 0 \tag{3.4.44}$$

$$\Big(p^2\,\mathbf{Z}_L^{-1} - \mathbf{M}\mathbf{Z}_R\mathbf{M}\Big)\xi^L(p) \;=\; 0\,, \tag{3.4.45}$$

using (3.4.19, 3.4.29) we find that the rotated doublets

$$\mathcal{U}[\theta_R(p^2)]\begin{pmatrix}\xi^R_\mu(p)\\ \xi^R_e(p)\end{pmatrix} = \begin{pmatrix}\varphi^R_1(p)\\ \varphi^R_2(p)\end{pmatrix} \;\;;\;\; \mathcal{U}[\theta_L(p^2)]\begin{pmatrix}\xi^L_\mu(p)\\ \xi^L_e(p)\end{pmatrix} = \begin{pmatrix}\varphi^L_1(p)\\ \varphi^L_2(p)\end{pmatrix} \tag{3.4.46}$$

obey the following equation

$$\begin{pmatrix} Q^R_\mu(p^2) - \varrho_R(p^2) & 0\\ 0 & Q^R_e(p^2) + \varrho_R(p^2)\end{pmatrix}\begin{pmatrix}\varphi^R_1(p)\\ \varphi^R_2(p)\end{pmatrix} = 0 \tag{3.4.47}$$

$$\begin{pmatrix} Q^L_\mu(p^2) - \varrho_L(p^2) & 0\\ 0 & Q^L_e(p^2) + \varrho_L(p^2)\end{pmatrix}\begin{pmatrix}\varphi^L_1(p)\\ \varphi^L_2(p)\end{pmatrix} = 0\,. \tag{3.4.48}$$

Neglecting perturbative renormalization of the μ, e masses, for $p^2 \simeq M^2_\mu$ the propagating modes correspond to $\varphi^{R,L}_1 \neq 0$; $\varphi^{R,L}_2 = 0$ and the mixing angles for R, L components are $\theta_{R,L}(M^2_\mu)$ respectively, with

$$\theta_R(M^2_\mu) \simeq \frac{M_e}{M_\mu}\, z^f_{L,\mu e}(M^2_\mu) \;\;;\;\; \theta_L(M^2_\mu) \simeq z^f_{L,\mu e}(M^2_\mu) \tag{3.4.49}$$

defining the μ-like propagating modes

$$\begin{pmatrix}\xi^L_\mu(p)\\ \xi^L_e(p)\end{pmatrix} = \varphi^L_1(p)\begin{pmatrix}\cos\theta_L(M^2_\mu)\\ \sin\theta_L(M^2_\mu)\end{pmatrix} \;\;;\;\; \begin{pmatrix}\xi^R_\mu(p)\\ \xi^R_e(p)\end{pmatrix} = \varphi^R_1(p)\begin{pmatrix}\cos\theta_R(M^2_\mu)\\ \sin\theta_R(M^2_\mu)\end{pmatrix} \tag{3.4.50}$$

Similarly for $p^2 \simeq M^2_e$ the propagating modes near the electron mass shell correspond to $\varphi^{R,L}_2 \neq 0$; $\varphi^{R,L}_1 = 0$ and the mixing angles for R, L components are $\theta_{R,L}(M^2_\mu)$ respectively, with

$$\theta_R(M^2_e) \simeq \frac{M_e}{M_\mu}\, z^f_{L,\mu e}(M^2_e) \;\;;\;\; \theta_L(M^2_e) \simeq \frac{M^2_e}{M^2_\mu}\, z^f_{L,\mu e}(M^2_e) \tag{3.4.51}$$

defining the relation between the flavor doublets and the propagating modes on the respective mass shells, namely

$$\begin{pmatrix} \xi^L_\mu(p) \\ \xi^L_e(p) \end{pmatrix} = \varphi^L_2(p) \begin{pmatrix} -\sin\theta_L(M_e^2) \\ \cos\theta_L(M_e^2) \end{pmatrix} \;\; ; \;\; \begin{pmatrix} \xi^R_\mu(p) \\ \xi^R_e(p) \end{pmatrix} = \varphi^R_2(p) \begin{pmatrix} -\sin\theta_R(M_e^2) \\ \cos\theta_R(M_e^2) \end{pmatrix} \tag{3.4.52}$$

The expressions (3.4.50,3.4.52) combined with (3.4.42,3.4.43) give a complete description of the propagating modes.

3.4.3 Alternative diagonalization procedure.

The quadratic part of the effective action in terms of the flavor doublet (3.4.1) and after renormalization is

$$\mathcal{L}_{eff} = \overline{\Psi}_R \not{p} \, \mathbf{Z}_R^{-1} \, \Psi_R + \overline{\Psi}_L \not{p} \, \mathbf{Z}_L^{-1} \, \Psi_L - \overline{\Psi}_R \mathbf{M} \Psi_L - \overline{\Psi}_L \mathbf{M} \Psi_R \tag{3.4.53}$$

In this expression $\mathbf{Z}_{R,L}$ are finite because the renormalization counterterms cancelled the divergent parts. These finite wavefunction renormalization *matrices* can be absorbed into a finite but four-momentum dependent renormalization of the Dirac fields, so that the kinetic terms are canonical, namely

$$\overline{\Psi}_{R,L}(p) = \overline{\eta}_{R,L}(p) \sqrt{\mathbf{Z}_{R,L}(p)} \;\; ; \;\; \Psi_{R,L}(p) = \sqrt{\mathbf{Z}_{R,L}(p)} \; \eta_{R,L}(p) \, , \tag{3.4.54}$$

leading to

$$\mathcal{L}_{eff} = \overline{\eta}_R \not{p} \, \eta_R + \overline{\eta}_L \not{p} \, \eta_L - \overline{\eta}_R \mathcal{M}(p) \eta_L - \overline{\eta}_L \, \mathcal{M}^\dagger(p) \, \eta_R \, , \tag{3.4.55}$$

we emphasize that because $\mathbf{Z}_L$ features off-diagonal terms, the above transformation is *not only* a simple rescaling but also a *mixing* between the μ, e fields.

The mass matrices

$$\mathcal{M}(p) = \sqrt{\mathbf{Z}_R(p)} \; \mathbf{M} \; \sqrt{\mathbf{Z}_L(p)} \;\; ; \;\; \mathcal{M}^\dagger(p) = \sqrt{\mathbf{Z}_L(p)} \; \mathbf{M} \; \sqrt{\mathbf{Z}_R(p)} \tag{3.4.56}$$

feature off diagonal terms from $\mathbf{Z}_L$ and are *momentum dependent*. They can be diagonalized by *biunitary transformations*, namely introducing the *unitary* matrices $\mathcal{V}_{R,L}$ as

$$\eta_{R,L} = \mathcal{V}_{R,L} \, \Phi_{R,L} \; ; \; \overline{\eta}_{R,L} = \overline{\Phi}_{R,L} \, \mathcal{V}^\dagger_{R,L} \tag{3.4.57}$$

these matrices are momentum dependent and diagonalize the mass matrices,

$$\mathcal{V}_R^\dagger \, \mathcal{M} \, \mathcal{V}_L = M_d \quad ; \quad \mathcal{V}_L^\dagger \, \mathcal{M}^\dagger \, \mathcal{V}_R = M_d \tag{3.4.58}$$

where M_d is a diagonal but momentum dependent "mass" matrix. It is straightforward to prove that

$$\mathcal{V}_R^\dagger \, \mathcal{M} \, \mathcal{M}^\dagger \, \mathcal{V}_R = M_d^2 = \mathcal{V}_L^\dagger \, \mathcal{M}^\dagger \, \mathcal{M} \, \mathcal{V}_L \, . \tag{3.4.59}$$

Projecting the Dirac equation obtained from the effective action (3.4.53) onto left and right handed components we find

$$\big[p^2 - \mathcal{M}^\dagger \, \mathcal{M}\big]\eta_L = 0 \quad ; \quad \big[p^2 - \mathcal{M} \, \mathcal{M}^\dagger\big]\eta_R = 0 \tag{3.4.60}$$

which are diagonalized by the unitary transformation (3.4.57) with the property (3.4.59). Obviously the position of the mass shells which are determined by the zeroes of the determinant of the operators in the brackets are the same as those obtained from the un-scaled Dirac equations (3.4.44,3.4.45) a result that is straightforwardly confirmed.

The μ-like and e-like eigenvectores are

$$\Phi_R(p) = \varphi_1^R(p) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ for } p^2 = M_\mu^2 + \cdots \quad ; \quad \Phi_R(p) = \varphi_2^R(p) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ for } p^2 = M_e^2 + \cdots \tag{3.4.61}$$

where the dots stand for the radiative corrections to the masses. After straightforward algebra we find to leading order

$$\mathcal{V}_R[\delta_R] = \begin{pmatrix} \cos\delta_R & -\sin\delta_R \\ \sin\delta_R & \cos\delta_R \end{pmatrix} \quad ; \quad \delta_R(p) \simeq \frac{M_\mu M_e \, z^f_{L,\mu e}(p^2)}{M_\mu^2 - M_e^2} \simeq \frac{M_e}{M_\mu} \, z^f_{L,\mu e}(p^2) \tag{3.4.62}$$

which is exactly the same as the rotation angle for the right handed component $\theta_R(p)$ given by eqn. (3.4.17) when evaluated on the mass shells $p^2 \simeq M_\mu^2$; $p^2 \simeq M_e^2$ respectively.

For the left handed component the μ-like and e-like eigenvectores are

$$\Phi_L(p) = \varphi_1^L(p) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ for } p^2 = M_\mu^2 + \cdots \quad ; \quad \Phi_R(p) = \varphi_2^L(p) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ for } p^2 = M_e^2 + \cdots \tag{3.4.63}$$

and again to leading order we find

$$\mathcal{V}_L[\delta_L] = \begin{pmatrix} \cos\delta_L & -\sin\delta_L \\ \sin\delta_L & \cos\delta_L \end{pmatrix} \quad ; \quad \delta_L(p) \simeq \frac{1}{2}\left(\frac{M_\mu^2 + M_e^2}{M_\mu^2 - M_e^2}\right) z^f_{L,\mu e}(p^2)\,. \tag{3.4.64}$$

We are now in position to reverse the re-scaling and unitary transformation to obtain the relation between the original μ, e fields and the fields that diagonalize the effective action, namely from (3.4.54), and (3.4.57) it follows that

$$\Psi_{R,L} = \begin{pmatrix} \xi^{R,L}_\mu \\ \xi^{R,L}_e \end{pmatrix} = \sqrt{\mathbf{Z}_{R,L}}\; \mathcal{V}_{R,L}\, \Phi_{R,L}\,. \tag{3.4.65}$$

Since $\mathbf{Z}_R$ is diagonal, we find to leading order

$$\begin{pmatrix} \xi^R_\mu(p) \\ \xi^R_e(p) \end{pmatrix}_{p^2\simeq M_\mu^2} \simeq \varphi^R_1(p) \begin{pmatrix} 1 \\ \theta_R(M_\mu^2) \end{pmatrix} \quad ; \quad \begin{pmatrix} \xi^R_\mu(p) \\ \xi^R_e(p) \end{pmatrix}_{p^2\simeq M_e^2} \simeq \varphi^R_2(p) \begin{pmatrix} -\theta_R(M_e^2) \\ 1 \end{pmatrix}. \tag{3.4.66}$$

The matrix $\mathbf{Z}_L$ is off-diagonal so that to leading order it follows that

$$\sqrt{\mathbf{Z}_L(p)} = \begin{pmatrix} 1+\cdots & \frac{1}{2}\, z^f_{L,\mu e}(p^2) \\ \frac{1}{2}\, z^f_{L,\mu e}(p^2) & 1+\cdots \end{pmatrix} \tag{3.4.67}$$

combining this result with (3.4.64) we find to leading order

$$\begin{pmatrix} \xi^L_\mu(p) \\ \xi^L_e(p) \end{pmatrix}_{p^2\simeq M_\mu^2} \simeq \varphi^R_1(p) \begin{pmatrix} 1 \\ \theta_L(M_\mu^2) \end{pmatrix} \quad ; \quad \begin{pmatrix} \xi^L_\mu(p) \\ \xi^L_e(p) \end{pmatrix}_{p^2\simeq M_e^2} \simeq \varphi^L_2(p) \begin{pmatrix} -\theta_L(M_e^2) \\ 1 \end{pmatrix} \tag{3.4.68}$$

where $\theta_L(p^2)$ is given by eqn. (3.4.35) with $p^2 \simeq M_\mu^2\,,\, M_e^2$ respectively.

Thus we have confirmed that the alternative diagonalization procedure with rescaling the fields and diagonalizing the resulting mass matrices with bi-unitary transformations yield the same result as the direct procedure described in the previous sections, thereby establishing that the results obtained above are robust.

3.5 RELATION TO LEPTON FLAVOR VIOLATING PROCESSES:

Charged lepton *mixing* via intermediate states of charged vector bosons and neutrino mass eigenstates are *directly* related to lepton flavor violating processes. An important process that is currently the focus of experimental searches[136, 137] and a recent proposal[138] is the decay $\mu \to e\,\gamma$ which is mediated by neutrino mass eigenstates[143, 144, 145, 146] and the importance of heavy sterile neutrinos in this process has been highlighted in ref.[147]. However, to the best of our knowledge the relationship between this process and a *mixed* $\mu - e$ propagator has not yet been explored. Such relationship is best understood in terms of the *three-loop* muon self-energy diagram in fig. (19-(a)), the Cutkosky cut along the intermediate state of the electron and photon yields the imaginary part of the muon propagator on its mass shell, and determines the decay rate $\mu \to e\gamma$, this is depicted in fig. (19-(b)).

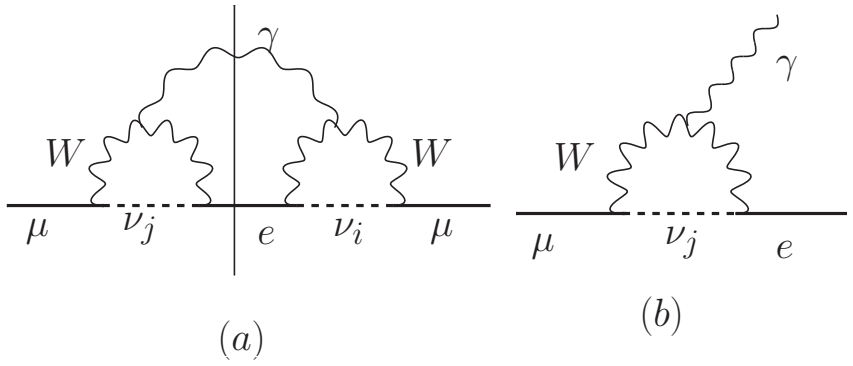


Figure 19: Lepton flavor violation: fig. (a): three loop contribution to $\Sigma_{\mu\mu}$ the Cutkosky cut through the photon and electron intermediate state yields the imaginary part describing the flavor violating decay $\mu \to e\gamma$ of fig.(b).

However, the self-energy diagram (19-(a)) is only *one diagonal component* of the full $\mu - e$ self-energy, the corresponding three loop diagram for the off-diagonal component is shown in fig. 20.

Because of the different external particles, a Cutkosky cut of this diagram through the photon and electron internal lines cannot be interpreted as a decay rate. However, this analysis clearly indicates the relationship between $\mu \to e\gamma$, a distinct indicator of lepton

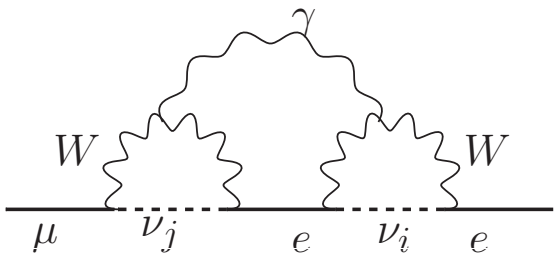


Figure 20: Three loop contribution to $\Sigma_{\mu e}$ which is the off-diagonal counterpart of fig. (19(a).

flavor violation, and charged lepton mixing in self-energy diagrams, both a direct consequence of neutrino mixing.

We note that whereas the branching ratio for $\mu \to e\gamma$ is $\propto G_F\alpha|\sum_j U_{\mu j}U_{je}|^2 m_j^2$ we find that the one-loop mixing angles are momentum dependent, different for different chiralities and the largest angle for on-shell states corresponds to the negative chirality muon-like combination, in which case the angle is of order $G_F\sum_j U_{\mu j}U^*_{je}m_j^2$.

***Possible* other contributions:** The diagram in fig. (19-(b)) suggests that $\mu - e$ mixing *may* lead to further contributions. Consider the flavor blind electromagnetic vertices of μ and e, *if* the mixing angles were momentum independent, unitarity of the transformation would entail a GIM cancellation between off-diagonal terms in the electromagnetic vertices, just as for neutral currents. However, muon-like and electron-like mass shells feature very different mixing angles which *suggests* that off diagonal contributions arising from replacing the μ and e fields in the electromagnetic vertices by the correct propagating states would *not cancel out* because of different mixing angles. This can be seen from the relation between the propagating states and the μ, e states given by eqns. (3.4.50,3.4.52), writing

$$\psi_\mu = \cos\theta_1\varphi_1 - \sin\theta_2\varphi_2 \quad ; \quad \psi_e = \cos\theta_2\varphi_2 + \sin\theta_1\varphi_1 \tag{3.5.1}$$

respectively for positive and negative chirality components with the respective angles $\theta_{1L} = \theta_L(M_\mu^2); \theta_{2L} = \theta_L(M_e^2)$ etc., it follows that the electromagnetic vertices feature a mixed term

of the form

$$\propto \overline{\varphi}_{2L}\gamma^{\mu}A_{\mu}\varphi_{1L}(\theta_{1L}-\theta_{2L})+L\to R\,, \tag{3.5.2}$$

where the right handed angles are very different from the left handed counterparts. If the mixing angle(s) were momentum independent $\theta_1 = \theta_2$ and this term would vanish in a manner similar to the GIM mechanism. Furthermore the *difference* $\theta_{1L}-\theta_{2L}$ is insensitive to the choice of the local renormalization counterterms. Therefore mixing with *momentum and chirality dependent mixing angles* suggests that the contribution to $\mu \to e\gamma$ from the vertex (3.5.2) depicted in fig. (21) becomes possible.

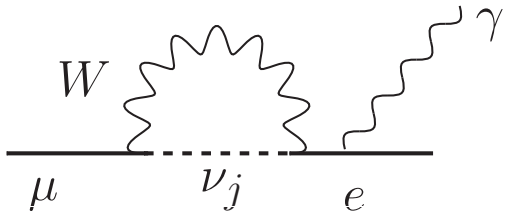


Figure 21: Further contribution to $\Sigma_{\mu e}$ from $\mu - e$ mixing.

This contribution differs from that of fig. (19-(b)) in two major aspects: i) rather than an extra W propagator, it features an electron propagator in the intermediate state, which would suggest a large enhancement with respect to the usual contribution, ii) a very small mixing angle which suppresses the enhancement from the electron propagator in the intermediate state. Thus a detail study of both effects and their impact is required for a firmer assessment.

This argument, however, needs to be scrutinized further by analyzing the imaginary part of the propagators keeping both the diagonal electromagnetic contribution as well as the off diagonal charged current contribution. Upon the diagonalization of the propagator there is an interference between the diagonal and the off-diagonal terms, this can be gleaned from (3.4.14,3.4.25). The imaginary part of the propagator evaluated at the mass shell of the muon-like propagating mode, namely $p^2 \simeq M_{\mu}^2$ would yield the contribution to the diagram of fig. (21) from the interference between the diagonal electromagnetic contribution which features an imaginary part for $p^2 > M_e^2$ and the off-diagonal charged current contribution. Since this is a contribution to the self energy of higher order than the ones considered

here, a firmer assessment of this new contribution merits further study and will be reported elsewhere.

3.6 SUMMARY AND DISCUSSION:

We summarize and clarify some of the main results obtained above.

- **Mixing:** In this study mixing refers to the fact that the flavor eigenstates of charged leptons, $\mu; e$ are *not the propagating states*. This is a consequence of self-energy corrections that are *off-diagonal* in the flavor basis as a consequence of intermediate states with neutrino mass eigenstates that connect the flavor states. As is standard in quantum field theory, the propagating modes correspond to the poles of the full propagator, because of the off-diagonal self-energy contributions these propagators become an off diagonal matrix in flavor space, whose diagonalization yields the correct propagating modes. We offered two complementary methods to understand the mixing and diagonalization: (i) a direct diagonalization of the propagator matrix including the one-loop self energy which features the off-diagonal terms, (ii) a diagonalization of the effective action by first rescaling the fields to a canonical form followed by a bi-unitary transformation to diagonalize the mass terms. Both approaches yield the same result: mixing angles that depend on the corresponding mass shells and different for right and left-handed components, these are given by (3.4.49,3.4.51) which are also obtained via the procedure of rescaling the fields to a canonical form, diagonalizing the mass matrices by a bi-unitary transformation and re-scaling back to find the relation between the original flavor eigenstates and the propagating eigenstates yielding the same mixing angles (see the discussion below eqns. (3.4.62,3.4.68). As described with both methods, the transformations necessary to relate the flavor and propagating eigenstates are manifestly *non-local* which is reflected on the different mixing angles on the different mass shells.

  The mixing angles are GIM suppressed favoring heavier neutrinos in the intermediate state and *momentum and chirality dependent*. This means that off-shell processes nec-

essarily mix charged leptons with *virtuality and chirality dependent mixing angles*. For $p^2 \ll M_W^2$ and assuming that the heaviest sterile neutrinos feature masses $\ll M_W^2$, from eqn. (3.3.19) we find the positive and negative chirality mixing angles for $\mu - e$ mixing

$$\theta_R \simeq \frac{G_F}{4\pi^2}\frac{M_e}{M_\mu}\sum_j U_{\mu j}U_{je}^* m_j^2 \quad ; \quad \theta_L(p^2) \simeq \frac{G_F}{4\pi^2}\frac{p^2}{M_\mu^2}\sum_j U_{\mu j}U_{je}^* m_j^2 \tag{3.6.1}$$

thus the mixing angles are dominated by the heaviest generation of neutrinos, and the *difference* of mixing angles at the different mass shells is insensitive to the choice of local renormalization counterterms. In particular *if* heavy sterile neutrinos do exist, these new degrees of freedom will yield the largest contribution to charged lepton mixing. Considering for example that there is only one generation of heavy sterile neutrinos with mass M_S, and *assuming* that $U_{\mu i} \simeq U_{ei}$, the recent results from the PIENU collaboration at TRIUMF[126] reporting an upper limit $|U_{ei}|^2 \leq 10^{-8}$ $(90\%\, C.L.)$ in the neutrino mass region $60 - 129\,\mathrm{MeV}/c^2$ allows us to estimate an upper bound for the negative chirality mixing angle near the μ mass shell,

$$\theta_L(p^2 \simeq M_\mu^2) - \theta_L(p^2 \simeq M_e^2) \leq 10^{-14}\left(\frac{M_S}{100\,\mathrm{MeV}}\right)^2 . \tag{3.6.2}$$

Oscillations are manifest in the off diagonal density matrix elements in the *flavor basis*. These, however, average out on unobservable small time scales thus coherence (off-diagonal density matrix elements in the flavor basis) is suppressed by these rapid oscillations and is not experimentally relevant.

- **Renormalization:** The off-diagonal component of the self-energy (in the $\mu - e$ basis), features ultraviolet divergences, which are regularized in dimensional regularization consistently with the underlying gauge symmetry. The renormalization counterterm has been chosen in the $\overline{MS}$ scheme as is commonly done. The fact that the renormalized Lagrangian requires an off-diagonal counterterm is again a consequence of the fact that intermediate states with neutrino mass eigenstates mix the flavor fields $\mu - e$. However, the counterterm in the renormalized Lagrangian is *local* and cannot completely remove the mixing between the flavor fields, this is manifest in the non-local and finite contribution to the off-diagonal self-energy given by eqn. (3.3.9) for the short distance contribution and (3.3.14) for the long-distance contribution. These *finite* contributions

are momentum dependent and feature absorptive cuts above the two particle threshold corresponding to the intermediate state of a charged vector boson and a neutrino mass eigenstate.

The momentum dependence leads to the different mixing angles on the mass shells as discussed in detail in the previous section, and the absorptive part gives rise to off-shell processes that involve the mixing of the flavor fields. In particular the *difference* between the mixing angles at the two mass shells is independent of the local counterterm which is also obviously irrelevant for the absorptive part.

- **Lepton flavor violation:** The relationship between the off-diagonal self energy and lepton flavor violating processes becomes manifest by explicitly comparing the Feynman diagram in fig. (18) for the self energy with $l_\alpha^- = \mu; l_\beta^- = e$ with that of the lowest order lepton flavor violating process $\mu \to e\gamma$ in fig. (19 -(b)): neglecting the photon line, the intermediate state of $W - \nu_j$ is the *same* as for the self-energy (18), namely: the mixing of flavors as a consequence of an off-diagonal self-energy in the $\mu - e$ basis has the same *physical origin* as the lepton-flavor violating process $\mu \to e\gamma$. The direct relationship between the off-diagonal self-energy and $\mu \to e\gamma$ is shown explicitly in figs. (19,20). Diagram (19-(a)) is the $\mu - \mu$ (diagonal) part of the self-energy, its Cutkosky cut across the W-line yields the imaginary part describing the process $\mu \to e\gamma$ in (19-(b)). The *same* types of intermediate states yield the *off-diagonal* $\mu - e$ contribution to the self-energy, displayed in fig.(20) clearly indicating that the *physical origin* of the *mixing* of $\mu - e$ flavor fields is the *same* as the lepton flavor violating transitions $\mu \to e\gamma$. Upon writing the charged lepton fields in terms of the propagating modes in flavor diagonal vertices in the interaction Lagrangian, the momentum dependent field redefinition associated with the rescaling and bi-unitary transformation, namely the mixing, yields novel interaction vertices in terms of the propagating modes that depend on the *difference* of the mixing angles at the different mass shells, this difference is independent of the *local* renormalization counterterm. A simple example is the electromagnetic vertex which is flavor diagonal, upon writing it in terms of the propagating modes $\varphi_{1,2}$ it describes an interaction between these in terms of the difference between the mixing angles at the mass shells, see eqn. (3.5.2) that leads to potentially new observable contributions such

as that displayed in fig.(21) that merit further study.

3.7 CONCLUSIONS AND FURTHER QUESTIONS

In this article we studied charged lepton *oscillations* and *mixing*. The decay of pseudoscalar mesons leads to an entangled quantum state of neutrinos and charged leptons (we focused on π, K decay leading to μ, e). If the neutrinos are not observed, tracing over their degrees of freedom leads to a density matrix for the charged leptons whose off-diagonal elements in the flavor basis reveals *charged lepton oscillations*. While these oscillations decohere on unobservably small time scales $\lesssim 10^{-23}\, s$, we recognize that they originate in a common set of intermediate states for the charged leptons. This realization motivated us to study the mixed $\mu - e$ *self-energies* and we recognized that charged-current interactions lead to a dominant "short distance" contribution to $\mu - e$ mixing via W-exchange and an intermediate neutrino mass eigenstate, and a subdominant (by a large factor) "long distance" contribution to mixing via an intermediate state with a pseudoscalar meson and neutrino mass eigenstate. We include the leading contribution in the propagator matrix for the $\mu - e$ system focusing on the off-diagonal terms which imply $\mu - e$ mixing. We find that the mixing angles are *chirality and momentum dependent*, the chirality dependence is a consequence of $V - A$ charge current interactions. Diagonalizing the kinetic term and the mass matrix by bi-unitary transformations or alternatively diagonalizing the propagator, displays the poles which describe muon-like and electron-like propagating modes ("mass eigenstates") for which we find explicitly the wave functions, but the mixing angles evaluated on the respective mass shells (and chiralities) are very different. We find the positive and negative chirality momentum dependent mixing angles for $p^2 \ll M_W^2$ to be approximately given by

$$\theta_R \simeq \frac{G_F}{4\pi^2}\frac{M_e}{M_\mu}\sum_j U_{\mu j}U_{je}^*\, m_j^2 \quad ; \quad \theta_L(p^2) \simeq \frac{G_F}{4\pi^2}\frac{p^2}{M_\mu^2}\sum_j U_{\mu j}U_{je}^*\, m_j^2 \tag{3.7.1}$$

therefore dominated by the heaviest generation of sterile neutrinos. The *difference* of mixing angles at the different mass shells is independent of local renormalization counterterms. For

one (dominant) generation of massive sterile neutrinos with mass M_S, the recent results from the PIENU collaboration at TRIUMF[126], suggests

$$\theta_L(p^2 \simeq M_\mu^2) - \theta_L(p^2 \simeq M_e^2) \leq 10^{-14} \left(\frac{M_S}{100\,\mathrm{MeV}}\right)^2 . \tag{3.7.2}$$

Flavor diagonal interaction vertices feature novel interactions once written in terms of the fields associated with the propagating modes or mass eigenstates. In particular the electromagnetic vertex, yields an interaction between the muon-like and electron-like propagating modes which is another manifestation of lepton flavor violation. The (four) momentum dependence of the $\mu - e$ mixing angle *may* be the source of novel off-shell effects whose potential observational manifestation merits further study. We expect to report on ongoing study on these issues elsewhere.

We discussed the relationship between the lepton flavor violating decay $\mu \to e\gamma$, the focus of current searches[136, 137] and proposals[138], and charged lepton mixing, pointing out that a positive measurement of the former confirms the latter. Furthermore, we advance the possibility of further contributions to $\mu \to e\gamma$ arising from the fact that the $\mu - e$ mixing angle is momentum dependent and differs substantially on the mass shells of the propagating modes voiding a GIM mechanism for the electromagnetic vertices.

Furthermore, in order to present the main arguments in the simplest case, in this article we have not considered CP-violating phases in the mixing matrix elements $U_{\alpha j}$, including these phases merit further study since this aspect could indicate potentially rich CP-violating phenomena from the charged lepton sector *induced* by CP-violation from the neutrino sector which merits further and deeper study.

4.0 ENTANGLEMENT ENTROPY IN PARTICLE DECAY.

Based on: (ref. [198])

L. Lello, D. Boyanovsky, R. Holman, JHEP 2013 116

4.1 INTRODUCTION

Once described as the source of "spooky action at a distance" by Einstein, Podolsky and Rosen (EPR)[148], quantum entanglement has now come to be viewed as a resource to be exploited in a number of venues. It serves as the workhorse for quantum computations[149, 150, 151, 152, 153, 154, 155], and is at the heart of current efforts on quantum information. In condensed matter and quantum optics, the spontaneous decay of excited atomic states leads to quantum entangled states of photons and atoms or spin-qubits, which can then be implemented as platforms for quantum computing[150, 151] by transmitting the information stored in the quantum correlations of the entangled states. Furthermore, current experiments in high energy physics are beginning to exploit the quantum correlations of pairs of particles produced in meson decay to study various aspects of CP violation and time reversal invariance.

The Belle collaboration[156] has reported on remarkably precise measurements of (EPR) entanglement and correlations in $\Upsilon(4S) \to B^0\overline{B}^0$ decays via the analysis of the time dependence of the flavor asymmetry. The BaBar collaboration[157, 158] has reported on the first direct measurement of time reversal violation in the $B^0\overline{B}^0$ system from $\Upsilon(4S)$ decay at rest

by studying the correlations between the members of the entangled $B^0\overline{B}^0$ pairs combining "flavor tagging" with "CP tagging". The possibility has also been advanced[159, 160, 161] of using entanglement correlations to establish bounds on the $B_s - \overline{B}_s$ width difference and CP violating phases. Entanglement between the charged lepton and its associated neutrino in the decay of pseudoscalar mesons has been recently argued to play an important role in the coherence (and decoherence) aspects of neutrino oscillations[87, 106, 104, 102] with potentially important corrections in short baseline oscillation experiments[162].

In the experiments that explore CP and or T violation from the time evolution of entangled states of $B^0\overline{B}^0$ pairs the main role played by entanglement is that the information contained in the quantum correlations of the entangled states is used to carry out measurement or "tagging" on one or both of the members. This article will explore a complementary aspect of entanglement and correlations, namely what happens when one or more members of an entangled state cannot (or will not) be measured, so that some of the information contained in the original quantum correlations of the entangled state is lost.

Given a pure quantum system consisting of entangled subsystems, it may not be possible to measure the separate state of all of the subsystems (or while possible, we may opt *not* to measure them). We can then construct a *reduced* density matrix for the subsystem(s) we do measure by tracing over the allowed states of the unobserved subsystem(s). This then leads directly to the concept of entanglement entropy: this is the von Neumann entropy of the reduced density matrix. It reflects the loss of information that was originally present in the entangled state from the quantum correlations. This entanglement entropy has been the focus of several studies in statistical and quantum field theories where subsystems are spatially correlated across boundaries by tracing over the degrees of freedom of one part of the system[163, 164, 165, 166, 167, 168] and have been extended to the case including black holes[169, 170], particle production in time dependent backgrounds[171] and cosmological space times[172]. Momentum space entanglement and renormalization has been recently studied in ref.[173].

While entanglement entropy has been mostly studied within the context of a quantum system subdivided by space-like regions (see references above), in this article we study the

time evolution of the entanglement entropy in the ubiquitous case of particle decay. The construction of the relevant states in this case relies on the Wigner-Weisskopf theory of spontaneous emission[100, 204], which provides a *non-perturbative* method for obtaining the quantum state arising from spontaneous decay. The knowledge of the full quantum state can then be used to obtain the entanglement entropy contained in, for example, the photon-spin qubit correlations generated from the dynamics of spontaneous decay in solid state systems[174].

Recently this theory was generalized to relativistic quantum field theory to yield insight into the quantum states from particle decay in cosmology[175], and to describe potential decoherence effects in neutrino oscillations in short baseline experiments[102, 162].

Motivation and goals:

Motivated by experiments at Belle and Babar that take advantage of the quantum correlations in entangled $B^0\overline{B}^0$ pairs produced from the decay at rest of the $\Upsilon(4S)$ resonance and by proposals to study CP violating phases by exploiting correlations in entangled $B_s\overline{B}_s$ pairs, we focus on a *complementary* aspect of entanglement, namely quantifying the loss of information if one of the members of the entangled pair is not measured. Tracing over the degrees of freedom of one of the members of an entangled pairs leads to a *reduced* density matrix, from which we obtain the entanglement or Von-Neumann entropy which *could* potentially be a useful tool to infer correlations even when particles in the final state are unobserved.

Here we extend and generalize the Wigner-Weisskopf method discussed in[162, 175, 174] to describe particle decay in quantum field theory and apply it to the simple case of a bosonic parent particle decaying into two bosonic daughter particles to highlight the main consequences, although we argue that the results are general. We address the important aspect of unitarity and obtain the entanglement entropy by tracing over the degrees of freedom associated with an unobserved daughter particle. We also show how unitary time evolution yields an entanglement entropy that grows over the lifetime of the parent particle and saturates to the entropy of maximally entangled states. Furthermore we extend the treatment to *wave packets* and compare with the case wherein the quantum states are described as

plane waves, assess the corrections and suggest a potential way to experimentally measure the entanglement entropy from the decay of mesons at rest.

4.2 THE WIGNER-WEISSKOPF METHOD

Consider a system described by a total Hamiltonian that can be decomposed as $H = H_0 + H_i$, where H_0 is the free Hamiltonian and H_i is the interaction part. As usual, the time evolution of a state in the interaction picture is given by

$$i\frac{d}{dt}|\Psi(t)\rangle = H_I(t)|\Psi(t)\rangle \tag{4.2.1}$$

where

$$H_I(t) = e^{iH_0 t} H_i e^{-iH_0 t} \tag{4.2.2}$$

is the interaction Hamiltonian in the interaction picture. The formal solution of (4.2.1) is given by

$$|\Psi(t)\rangle = U(t,t_0)|\Psi(t_0)\rangle \quad ; \quad U(t,t_0) = T(e^{-i\int_{t_0}^{t} H_I(t')dt'}) \tag{4.2.3}$$

Expanding the state $|\Psi(t)\rangle$ in the basis of eigenstates of H_0 we have

$$|\Psi(t)\rangle = \sum_n C_n(t)|n\rangle\,, \tag{4.2.4}$$

where $H_0|n\rangle = E_n|n\rangle$. From eq.(4.2.1), upon expanding in basis states $|n\rangle$, it follows that

$$\frac{dC_n(t)}{dt} = -i\sum_m \langle n|H_I(t)|m\rangle C_m(t)\,. \tag{4.2.5}$$

This is an infinite set of differential equations that can be solved hierarchically leading in general to integro-differential equations at any given order in the perturbative expansion. Progress can be made by considering the evolution of the states that are coupled at a given order in perturbation theory and solving the coupled equations for these, thereby truncating the hierarchy at a given order in the perturbative expansion.

For the case of interest here, consider an initial value problem in which the system is prepared at an initial time $t = 0$ in a state $|A\rangle$, so that $C_A(0) = 1,\ C_{n\neq A} = 0$ and that the Hamiltonian H_I couples these states to a set of states $|\kappa\rangle$. Then, we can close the hierarchy at second order in the interaction by keeping only the coupling of the states $|A\rangle \leftrightarrow |\kappa\rangle$, i.e.

$$\frac{d}{dt}\, C_A(t) = -i \sum_{\kappa} \langle A|H_I(t)|\kappa\rangle\, C_\kappa(t) \tag{4.2.6}$$

$$\frac{d}{dt}\, C_\kappa(t) = -i \langle \kappa|H_I(t)|A\rangle\, C_A(t)\,. \tag{4.2.7}$$

Using the initial conditions we obtain

$$C_\kappa(t) = -i \int_0^t dt' \langle \kappa|H_I(t')|A\rangle C_A(t')\,, \tag{4.2.8}$$

which when inserted into (4.2.6) leads to

$$\frac{d}{dt} C_A(t) = -\int_0^t dt' \sum_{\kappa} \langle A|H_I(t)|\kappa\rangle\langle \kappa|H_I(t')|A\rangle C_A(t') = -\int_0^t dt' \Sigma_A(t-t') C_A(t')\,, \tag{4.2.9}$$

where the second order self energy has been introduced

$$\Sigma_A(t-t') \equiv \sum_{\kappa} \langle A|H_I(t)|\kappa\rangle\langle \kappa|H_I(t')|A\rangle = \sum_{\kappa} \left|\langle A|H_I(0)|\kappa\rangle\right|^2 e^{i(E_A - E_\kappa)(t-t')}\,. \tag{4.2.10}$$

Higher order corrections can be included by enlarging the hierarchy, i.e. by considering the equations that couple the states κ to other states κ' via the Hamiltonian. The coefficients for the states κ' can be obtained by integration and can be inserted back in the equations for the coefficients C_κ (which already include their coupling to $|A\rangle$). Then formally integrating the equation and inserting the results back into the equation for $|A\rangle$ generates higher order corrections to the self-energy Σ_A. Finally, solving for the time evolution of C_A allows us to obtain the time evolution of the other coefficients.

As seen by the procedure described above, the Wigner-Weisskopf approach can be used to construct an approximate version of the quantum state in the presence of interactions.

However, what is not altogether obvious is that the truncation of states used to construct the state gives rise to a state whose time evolution is *unitary*. This will be extremely important in the sequel since we will want to follow that time evolution of the entanglement entropy this state would provide after tracing out an unobserved subsystem as discussed in the introduction. We will need to be sure that there are no spurious effects in this evolution due to an approximation to the state.

The statement of unitary is one of conservation of probability. From the evolution equation (4.2.5) and its complex conjugate it follows that

$$\frac{d}{dt}\sum_n |C_n(t)|^2 = -i\sum_{m,n}\left[C_m(t)C_n^*(t)\langle n|H_I(t)|m\rangle - C_n(t)C_m^*(t)\langle m|H_I(t)|n\rangle\right] = 0\,, \quad (4.2.11)$$

as can be seen by relabeling $m \leftrightarrow n$ in the second term. Therefore $\sum_n |C_n(t)|^2 =$ constant. Now this is an *exact* result; the question is whether and how is it fulfilled in the Wigner-Weisskopf approximation obtained by truncating the hierarchy to the set of equations (4.2.6, 4.2.7).

Using eqs.(4.2.8, 4.2.10) consider

$$\sum_\kappa |C_\kappa(t)|^2 = \int_0^t dt_1 C_A^*(t_1)\int_0^t dt_2 \Sigma_A(t_1,t_2)C_A(t_2). \quad (4.2.12)$$

Inserting $1 = \Theta(t_1-t_2)+\Theta(t_2-t_1)$ in the time integrals it follows that

$$\begin{aligned}\sum_\kappa |C_\kappa(t)|^2 &= \int_0^t dt_1 C_A^*(t_1)\int_0^{t_1} dt_2 \Sigma_A(t_1,t_2)C_A(t_2) \\ &+ \int_0^t dt_2 C_A(t_2)\int_0^{t_2} dt_1 \Sigma_A(t_1,t_2)C_A^*(t_1) \end{aligned} \quad (4.2.13)$$

so that using $\Sigma_A(t_1,t_2) = \Sigma_A^*(t_2,t_1)$ relabelling $t_1 \leftrightarrow t_2$ in the second line of (4.2.13) and using (4.2.9) we find

$$\begin{aligned}&\sum_\kappa |C_\kappa(t)|^2 = -\int_0^t dt_1\Big[C_A^*(t_1)\frac{d}{dt_1}C_A(t_1) + C_A(t_1)\frac{d}{dt_1}C_A^*(t_1)\Big] = -\int_0^t dt_1\frac{d}{dt_1}|C_A(t)|^2 \\ &= 1-|C_A(t)|^2 \end{aligned} \quad (4.2.14)$$

where we have used the initial condition $C_A(0) = 1$. This is the statement of unitary time evolution, namely

$$|C_A(t)|^2 + \sum_\kappa |C_\kappa(t)|^2 = |C_A(0)|^2 \tag{4.2.15}$$

This is an important result. As we will see below, standard perturbative calculations of this state would *not* yield a state that evolved unitarily. The Wigner-Weisskopf state involves a non-perturbative dressing up of the state and despite its approximate nature, this dressing up captures the physics to a sufficient extent to guarantee unitary time evolution.

4.3 REDUCED DENSITY MATRIX AND ENTANGLEMENT ENTROPY.

Now we are ready to turn to the problem we really want to consider: the state that appears after the decay of a parent particle Φ into two daughters χ, ψ. For simplicity and to highlight the main concepts we treat all fields as bosonic massive fields with masses m_Φ and m_χ, m_ψ respectively. We consider a typical interaction vertex described by the interaction Hamiltonian

$$H_I = g \int d^3x \, \Phi(\vec{x}) \, \chi(\vec{x}) \, \psi(\vec{x}) \, . \tag{4.3.1}$$

We quantize the fields in a volume V, so that in the interaction picture they can be written as:

$$\varphi(\vec{x}, t) = \sum_{\vec{k}} \frac{1}{\sqrt{2E_k V}} \left[a_{\vec{k}} \, e^{-iE_k t} \, e^{i\vec{k}\cdot\vec{x}} + a^\dagger_{\vec{k}} \, e^{iE_k t} \, e^{-i\vec{k}\cdot\vec{x}} \right] \quad ; \quad \varphi = \Phi, \chi, \psi \, . \tag{4.3.2}$$

For the case of interest here, namely the decay process $\Phi \to \chi \, \psi$ we consider that the initial state is a single particle Φ at rest, therefore the initial condition is $C_\Phi(\vec{k} = 0; t = 0) = 1$, $C_\kappa(t = 0) = 0$ for $|\kappa\rangle \neq |1^\Phi_{\vec{0}}\rangle$. The interaction Hamiltonian (4.3.1) connects the initial state, $|1^\Phi_{\vec{0}}\rangle$ to the states, $|\kappa\rangle = |\chi_{-\vec{p}}\rangle \otimes |\psi_{\vec{p}}\rangle$. These states in turn are coupled back to $|1^\Phi_{\vec{0}}\rangle$ via H_I, these processes are depicted in fig.(22).

Thus to leading order in g we find from the intermediate states shown in fig.(22)

$$\Sigma_\Phi(t - t') = \sum_{\vec{p}} |\langle 1^\Phi_{\vec{0}} | \hat{H}_I(0) | \chi_{-\vec{p}}, \psi_{\vec{p}} \rangle|^2 e^{i(m_\Phi - E_\chi(p) - E_\psi(p))(t-t')} \, . \tag{4.3.3}$$

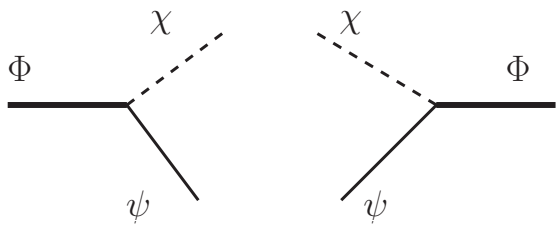


Figure 22: Transitions $|\Phi\rangle \leftrightarrow |\chi\rangle|\psi\rangle$ up to order g^2 that determine Σ_Φ.

The interaction Hamiltonian also connects a single Φ-particle state to an intermediate state with three other particles and this state back to the single Φ particle state yielding a disconnected contribution to the self energy depiced in fig.(23). This contribution is just a renormalization of the vacuum energy and only contributes to an overall phase that multiplies the single particle Φ state and will be neglected in the following analysis. For a more detailed discussion of this contribution see ref.[175].

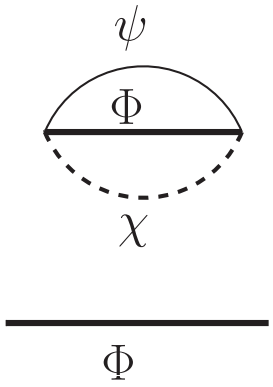


Figure 23: Order g^2 correction to the vacuum energy. Yields an overall phase for the quantum state $|1^\Phi\rangle$ [175].

At higher order in g there are higher order contributions to the Φ self energy from other multiparticle states. However, we will only be considering states connected to $|1^\Phi\rangle$ to first order in perturbation theory. As discussed in detail in refs.[162, 175] the Wigner-Weisskopf method provides a *non-perturbative resummation* of self-energies in *real time* akin to the

dynamical renormalization group[176]. The self-energy eq.(4.3.3) is recognized as the one-loop retarded self energy of the field Φ with the $|\chi\rangle|\psi\rangle$ intermediate state[162, 175].

Solving eq.(4.2.9) produces a solution for the time evolution of the Φ amplitude. We can use the solution for $C_\Phi(t)$ to obtain an expression for the amplitudes $C_\kappa(t)$ which allows for computation of the probability of occupying a particular state at any given time. We may solve eq.(4.2.9) either via Laplace transform, or in the case of weak coupling, a derivative expansion which yields the same result at long times ($t \gg 1/m_\Phi$). In ref.[175] the equivalence between the two approaches is discussed in detail. Here, we follow the latter method which is the original Wigner-Weisskopf approximation[100, 204].

We begin by defining the quantity

$$W_0(t,t') = \int_0^{t'} dt'' \Sigma_\Phi(t-t'') \tag{4.3.4}$$

so that

$$\frac{d}{dt'} W_0(t,t') = \Sigma_\Phi(t-t') \quad , \quad W_0(t,0) = 0 \tag{4.3.5}$$

Integrating eq.(4.2.9) by parts yields

$$\frac{d}{dt} C_\Phi(t) = -\int_0^t dt' \Sigma_\Phi(t-t')\ C_\Phi(t') = -W_0(t,t)\ C_\Phi(t) + \int_0^t dt' W_0(t,t') \frac{d}{dt'} C_\Phi(t'). \tag{4.3.6}$$

The first is term second order in H_I whereas the second term is of fourth order in H_I and will be neglected. This approximation is equivalent to the Dyson resummation of the one-loop self energy diagrams. Thus to leading order, eq.(4.2.9) becomes

$$\frac{d}{dt} C_\Phi(t) + W_0(t,t) C_\Phi(t) = 0\,, \tag{4.3.7}$$

where

$$W_0(t,t) = \int_0^t dt' \Sigma_\Phi(t-t') = \int_0^t dt' \sum_{\vec{p}} |\langle 1^\Phi_{\vec{0}} | \hat{H}_I(0) | \chi_{-\vec{p}}, \psi_{\vec{p}} \rangle|^2 e^{i(m_\Phi - E_\chi(p) - E_\psi(p))(t-t')} \tag{4.3.8}$$

Inserting a convergence factor and taking the limit $t \to \infty$ consistently with the Wigner-Weisskopf approximation, we find[1]

$$W_0(t,t) = \lim_{\epsilon \to 0^+} i \sum_{\kappa} \frac{|\langle 1_{\vec{0}}^{\Phi} | \hat{H}_I(0) | \chi_{-\vec{p}}, \psi_{\vec{p}} \rangle|^2}{m_\Phi - E_\chi(p) - E_\psi(p) + i\epsilon} = i\Delta E_\Phi + \frac{\Gamma}{2} \tag{4.3.9}$$

where

$$\Delta E_\Phi \equiv \mathcal{P} \sum_{\vec{p}} \frac{|\langle 1_{\vec{0}}^{\Phi} | \hat{H}_I(0) | \chi_{-\vec{p}}, \psi_{\vec{p}} \rangle|^2}{m_\Phi - E_\chi(p) - E_\psi(p)}\,, \tag{4.3.10}$$

is the second order shift in the energy which will be absorbed into a renormalization of the Φ mass and

$$\Gamma \equiv 2\pi \sum_{\vec{p}} |\langle 1_{\vec{0}}^{\Phi} | \hat{H}_I(0) | \chi_{-\vec{p}}, \psi_{\vec{p}} \rangle|^2 \delta(m_\Phi - E_\chi(p) - E_\psi(p)) \tag{4.3.11}$$

is the decay width as per Fermi's Golden rule. Therefore in this approximation, we arrive at

$$C_\Phi(t) = e^{-i\Delta E_\Phi t}\; e^{-\frac{\Gamma}{2}t}\,, \tag{4.3.12}$$

where we now consider a Φ with $\vec{k} = 0$ (decay at rest) and find

$$\mathcal{M}_\Phi(p) = \langle 1_{\vec{0}}^{\Phi} | \hat{H}_I(0) | \chi_{\vec{p}}\, \psi_{-\vec{p}} \rangle = \frac{g}{\sqrt{8V m_\Phi E_\chi(p) E_\psi(p)}} \tag{4.3.13}$$

leading to

$$\Gamma = 2\pi \sum_{p} |\mathcal{M}_\Phi(p)|^2 \delta\big(m_\Phi - E_\chi(p) - E_\psi(p)\big) = \frac{g^2\, p^*}{8\pi m_\Phi^2} \tag{4.3.14}$$

where

$$p^* = \frac{1}{2m_\Phi}\left[m_\Phi^4 + m_\chi^4 + m_\psi^4 - 2m_\Phi^2 m_\chi^2 - 2m_\Phi^2 m_\psi^2 - 2m_\chi^2 m_\psi^2 \right]^{1/2}. \tag{4.3.15}$$

Inserting (4.3.12) into (4.2.8) we find the quantum state

$$|\Psi(t)\rangle = e^{-i\Delta E_\Phi t}\; e^{-\frac{\Gamma}{2}t}\, |1_{\vec{0}}^{\Phi}; 0_\chi; 0_\psi\rangle + \sum_{\vec{p}} \mathcal{C}_{\chi\psi}(p\,;t)\, |\chi_{\vec{p}}\rangle\, |\psi_{-\vec{p}}|0_\Phi\rangle \tag{4.3.16}$$

[1]The long time limit in the Wigner-Weisskopf approximation is equivalent to the Breit-Wigner approximation of a resonant propagator[175] and holds for $t \gg 1/m_\Phi$.

where

$$\mathcal{C}_{\chi\psi}(p\,;t) = \mathcal{M}_{\Phi}(p)\,\frac{\left[1-e^{-i(m_{\Phi,R}-E_{\chi}(p)-E_{\psi}(p)-i\Gamma/2)\,t}\right]}{\left(E_{\chi}(p)+E_{\psi}(p)-m_{\Phi,R}+i\Gamma/2\right)} \tag{4.3.17}$$

and $m_{\Phi,R} = m_{\Phi} + \Delta E_{\Phi}$ is the renormalized mass of the Φ particle. In what follows we drop the subscript R and always refer to the renormalized mass.

At this stage we can make contact with the momentum entanglement discussion in ref.[173] and at the same time exhibit the true non-perturbative nature of the results above by considering the state obtained using naive perturbation theory.

Taking the initial state at $t = 0$ to be $|1^{\Phi}_{\vec{0}}\rangle$, then to leading order in g, the time evolved state is given by

$$|\Psi(t)\rangle = \left[1 - i\int_0^t e^{iH_0t'}H_I e^{-iH_0t'}dt' + \cdots\right]|1^{\Phi}_{\vec{0}}\rangle\,. \tag{4.3.18}$$

Introducing a resolution of the identity $1 = \sum_{\kappa}|\kappa\rangle\langle\kappa|$ we find

$$|\Psi(t)\rangle = |1^{\Phi}_{\vec{0}}\rangle + \sum_{\vec{p}}\left[\mathcal{M}_{\Phi}(p)\,\left(\frac{1-e^{-i(m_{\Phi}-E_{\chi}(p)-E_{\psi}(p))\,t}}{E_{\chi}(p)+E_{\psi}(p)-m_{\Phi}}\right)\right]|\chi_{-\vec{p}}\rangle\,|\psi_{\vec{p}}\rangle|0_{\Phi}\rangle + \cdots\,. \tag{4.3.19}$$

In this expression we have neglected a disconnected three particle intermediate state in which the initial $|1^{\Phi}_{\vec{0}}\rangle$ remains and the interaction creates an intermediate state with three particles. This contribution is truly perturbative, does not contain resonant denominators as (4.3.19) and corresponds to the disconnected diagram in fig.(23).

The probability of finding the daughter states is given by the familiar result (Fermi's Golden rule)

$$\mathcal{P}(t) = \sum_{\vec{p}}\left|\mathcal{M}_{\Phi}(p)\right|^2\left[\frac{\sin\left((m_{\Phi}-E_{\chi}(p)-E_{\psi}(p))\,t/2\right)}{(m_{\Phi}-E_{\chi}(p)-E_{\psi}(p))}\right]^2 = \Gamma t \tag{4.3.20}$$

where Γ is given by (4.3.14). This result is obviously only valid for $t \ll 1/\Gamma$. It is now clear that the generalized Wigner-Weisskopf method that yields the state (4.3.16) with the coefficients given by (4.3.17) is truly non-perturbative.

The momentum entanglement between the daughter particles is akin to that discussed perturbatively in ref.[173] with some important differences. In ref.[173] momentum entanglement was studied for the vacuum wave function. The corresponding contributions are truly

perturbative and do not feature the resonant denominators that lead to secular growth in time and are similar to the contributions that we neglect and are described by fig.(23).

The quantum state (4.3.16) describes an *entangled* state of the parent and daughter particles. The full (pure state) density matrix is given by

$$\rho(t) = |\Psi(t)\rangle\langle\Psi(t)| \tag{4.3.21}$$

and its trace is given by

$$\mathrm{Tr}\rho(t) = e^{-\Gamma t} + V\int \frac{d^3p}{(2\pi)^3}|\mathcal{C}_{\chi\psi}(p\,;t)|^2 \tag{4.3.22}$$

The momentum integral in (4.3.22) can be computed by changing variables to $\mathcal{E} = E_\chi(p) + E_\psi(p)$. In the narrow width limit $\Gamma \ll m_\Phi, m_\chi + m_\psi$, the integrand is sharply peaked at $\mathcal{E} = m_\Phi$ so that the lower limit can be consistently taken to $-\infty$ thus allowing the integral can be computed by contour integration. We find

$$V\int \frac{d^3p}{(2\pi)^3}|\mathcal{C}_{\chi\psi}(p\,;t)|^2 = 1 - e^{-\Gamma t}\,, \tag{4.3.23}$$

confirming that

$$\mathrm{Tr}\rho(t) = 1\,, \tag{4.3.24}$$

consistent with unitary time evolution and the unitarity relation (4.2.15). Furthermore the average number of ψ (or χ) particles is given by

$$n^\psi(p\,;t) = \langle\Psi(t)|a^\dagger_\psi(p)a_\psi(p)|\Psi(t)\rangle \equiv (2\pi)^3\frac{dN^\psi(t)}{d^3x d^3p} = |\mathcal{C}_{\chi\psi}(p\,;t)|^2\,, \tag{4.3.25}$$

thus the *total* number of ψ (or χ) particles is

$$N^\psi(t) = V\int \frac{d^3p}{(2\pi)^3}\, n^\psi(p\,;t) = 1 - e^{-\Gamma t}\,. \tag{4.3.26}$$

Tracing out one of the daughter particles, for example χ if it is unobservable, leads to a mixed state *reduced density matrix*

$$\rho_\psi(t) = \mathrm{Tr}_\chi\rho(t) = e^{-\Gamma t}\,|1^\Phi_{\vec{0}}\rangle\langle 1^\Phi_{\vec{0}}| + \sum_{\vec{p}}|\mathcal{C}_{\chi\psi}(p\,;t)|^2\,|\psi_{\vec{p}}\rangle\langle\psi_{\vec{p}}|\,. \tag{4.3.27}$$

The Von-Neumann entanglement entropy is therefore given by

$$S(t) = -n^{\Phi}(0,t)\,\ln[n^{\Phi}(0,t)] - \sum_{\vec{p}} n^{\psi}(p\,;t)\ln\left[n^{\psi}(p\,;t)\right]. \tag{4.3.28}$$

where $n^{\psi}(p)$ is given by (4.3.25) and $n^{\Phi}(0,t) = e^{-\Gamma t}$. Because in the narrow width limit $|\mathcal{C}_{\chi\psi}(p\,;t)|^2$ is a sharply peaked distribution, under integration with functions that vary smoothly near $p \simeq p^*$ it can be replaced by

$$n^{\psi}(p\,;t) = |\mathcal{C}_{\chi\psi}(p\,;t)|^2 \simeq \frac{2\pi^2 m_{\Phi}\left[1-e^{-\Gamma t}\right]}{V p^* E_{\chi}(p) E_{\psi}(p)}\;\frac{1}{2\pi}\;\frac{\Gamma}{(E_{\chi}(p)+E_{\psi}(p)-m_{\Phi})^2+(\Gamma/2)^2}\,. \tag{4.3.29}$$

Using this approximation we find

$$S(t) = \Gamma t e^{-\Gamma t} - \left[1-e^{-\Gamma t}\right]\ln\left[1-e^{-\Gamma t}\right] - \left[1-e^{-\Gamma t}\right]\ln\left[n^{\psi}(p^*\,;\infty)\right] \tag{4.3.30}$$

where

$$n^{\psi}(p^*\,;\infty) = \frac{4\pi\, m_{\Phi}}{V p^* E_{\chi}(p^*) E_{\psi}(p^*)\Gamma}\,. \tag{4.3.31}$$

and the asymptotic entanglement entropy is

$$S(\infty) = -\ln\left[\frac{4\pi\, m_{\Phi}}{V p^* E_{\chi}(p^*) E_{\psi}(p^*)\Gamma}\right]. \tag{4.3.32}$$

This result has the following interpretation. In the asymptotic limit $t \gg 1/\Gamma$, the entanglement (von Neumann) entropy approaches (minus) the logarithm of the available states. The decay of the parent particle produces entangled pairs in which each member features a very narrow distribution centered at p^* of width $\sim \Gamma$ and height $\sim 1/\Gamma$. The total area in momentum space yields $1/V$ since there is only one particle (of either type) produced in the volume V. Within the range of momenta centered at p^* and of width Γ all of the available single particle states have equal probability $\propto 1/V$, therefore these states are *maximally entangled* as Bell states. This observation becomes clearer recognizing that a typical quantum state that contributes to the sum in (4.3.16) is of the form

$$\mathcal{C}_{\chi\psi}(p^*,t)\Big[|\chi(\vec{p^*})\rangle\,|\psi(-\vec{p^*})\rangle + |\chi(-\vec{p^*})\rangle\,|\psi(\vec{p^*})\rangle\Big] \quad ; \quad \vec{p^*} = p^*\,\vec{n}\,, \tag{4.3.33}$$

where $\vec{n}$ is the direction of emission of either member of the pair. The quantum states with momenta $p^* - \Gamma/2 \leq p \leq p^* + \Gamma/2$ are represented with nearly the same probability $|\mathcal{C}_{\chi\psi}(p^*, t)|^2 \propto 1/V\Gamma$ in the sum. These states are Bell-type states and are maximally entangled, in fact these are similar, up to an overall normalization, to the entangled $B^0\overline{B^0}$ states resulting from the decay of the $\Upsilon(4S)$ resonance[156, 157, 158], but with the opposite relative sign because of charge conjugation. If the decaying particle has a short lifetime corresponding to a broad resonance, the emitted pairs will feature a distribution of momenta with probabilities $|\mathcal{C}_{\chi\psi}(p, t)|$ determined by the Lorentzian profile of the resonance.

As discussed above $n^{\psi}(p^*\,; \infty)$ (see eqn. (4.3.25)) is the asymptotic phase space density of the produced particle (either χ or ψ). The entanglement entropy vanishes at the initial time since the density matrix at $t = 0$ is a pure state and grows to its asymptotic value on a time scale $1/\Gamma$.

4.4 WAVE PACKETS

The treatment above described parent and daughter particles in terms of single particle plane waves, however in typical experiments the parent particle is produced as a wave packet with some localization length scale determined by the experimental setup. In this section we extend the treatment to a wave packet description. We use the discrete momentum representation in a quantization volume V.

Consider a particle of species $\alpha = \Phi, \chi, \psi$, Fock states describing single particle plane wave states of momentum $\vec{k}$, $|1^{\alpha}_{\vec{k}}\rangle$, are normalized such that

$$\langle 1^{\alpha}_{\vec{k}} | 1^{\alpha}_{\vec{k}'} \rangle = \delta_{\vec{k},\vec{k}'}\,. \tag{4.4.1}$$

Localized single particle states are constructed as linear superpositions

$$|\alpha; \vec{k}_0, \vec{x}_0\rangle = \sum_{\vec{k}} C_{\alpha}(\vec{k}; \vec{k}_0; \vec{x}_0)\, |1^{\alpha}_{\vec{k}}\rangle \tag{4.4.2}$$

where $C_\alpha(\vec{k};\vec{k}_0;\vec{x}_0)$ is the amplitude, normalized so that

$$\langle\alpha;\vec{k}_0,\vec{x}_0|\alpha;\vec{k}_0,\vec{x}_0\rangle = \sum_{\vec{k}} |C_\alpha(\vec{k};\vec{k}_0;\vec{x}_0)|^2 = 1\,. \tag{4.4.3}$$

The total number of particles in the volume V is

$$\langle\alpha;\vec{k}_0,\vec{x}_0|\sum_{\vec{k}} a^\dagger_{\alpha,\vec{k}} a_{\alpha,\vec{k}}|\alpha;\vec{k}_0,\vec{x}_0\rangle = 1\,. \tag{4.4.4}$$

For a monochromatic plane wave $C_\alpha(\vec{k};\vec{k}_0;\vec{x}_0) = \delta_{\vec{k},\vec{k}_0}$. The spatial wave function corresponding to the wave packet is given by

$$\Xi(\vec{x}) = \frac{1}{\sqrt{V}}\sum_{\vec{k}} C_\alpha(\vec{k};\vec{k}_0;\vec{x}_0)\, e^{-i\vec{k}\cdot\vec{x}}\,, \tag{4.4.5}$$

the normalization (4.4.3) implies

$$\int d^3x |\Xi(\vec{x})|^2 = 1\,. \tag{4.4.6}$$

For a monochromatic plane wave it follows that $\Xi(\vec{x})$ is a volume normalized plane wave.

The average momentum of the wave packet state is given by

$$\langle\alpha;\vec{k}_0,\vec{x}_0|\sum_{\vec{k}} \vec{k}\, a^\dagger_{\alpha,\vec{k}} a_{\alpha,\vec{k}}|\alpha;\vec{k}_0,\vec{x}_0\rangle = \sum_{\vec{k}} \vec{k}|C_\alpha(\vec{k};\vec{k}_0;\vec{x}_0)|^2 \tag{4.4.7}$$

where $a^\dagger_{\alpha,\vec{k}}; a_{\alpha,\vec{k}}$ are the creation and annihilation operators for the species α. Assuming that the distribution $C_\alpha(\vec{k};\vec{k}_0;\vec{x}_0)$ is isotropic in the rest frame of the wave-packet and that the average momentum is $\vec{k}_0$, it follows that

$$C_\alpha(\vec{k};\vec{k}_0;\vec{x}_0) = C_\alpha(\vec{k}-\vec{k}_0;\vec{x}_0)\,. \tag{4.4.8}$$

As a specific example we consider Gaussian wave packets,

$$C_\alpha(\vec{k}-\vec{k}_0;\vec{x}_0) = \left[\frac{8\,\pi^{\frac{3}{2}}}{\sigma^3\,V}\right]^{\frac{1}{2}} e^{-\frac{(\vec{k}-\vec{k}_0)^2}{2\sigma^2}}\; e^{i(\vec{k}-\vec{k}_0)\cdot\vec{x}_0}\,, \tag{4.4.9}$$

where σ is the localization in momentum space, the spatial wave function is

$$\Xi(\vec{x}) = \left[\frac{\sigma}{\sqrt{\pi}}\right]^{3/2} e^{-i\vec{k}_0\cdot\vec{x}}\, e^{-\frac{\sigma^2}{2}(\vec{x}-\vec{x}_0)^2}\,. \tag{4.4.10}$$

The spatial wave function is localized at $\vec{x}_0$ with localization length $1/\sigma$ and the momentum wave function is localized at $\vec{k}_0$ which is the average momentum in the wave packet and the momentum localization scale is σ. The plane wave limit is obtained by formally identifying $\sigma/\sqrt{\pi} \to 1/V^{1/3}$; $V \to \infty$.

4.4.1 Macroscopic localization and orthogonality:

In terms of these wave functions the overlap of two wave packets with different momenta localized at different spatial points is

$$\langle \alpha; \vec{q}_0; \vec{y}_0 | \alpha; \vec{k}_0; \vec{x}_0 \rangle = e^{-\frac{(\vec{k}_0 - \vec{q}_0)^2}{4\sigma^2}} \; e^{-\frac{\sigma^2}{4}(\vec{x}_0 - \vec{y}_0)^2} \; e^{-\frac{i}{2}(\vec{k}_0 - \vec{q}_0)\cdot(\vec{x}_0 - \vec{y}_0)} \, . \tag{4.4.11}$$

For a macroscopic localization length $1/\sigma$ the wavepackets are nearly orthogonal in momentum for values of the momentum of experimental relevance. For example, consider a localization length $\simeq 1\,\text{meter}$

$$1/\sigma \simeq 1 \text{ meter} \Rightarrow \sigma \simeq 2 \times 10^{-7}\,\text{eV} \tag{4.4.12}$$

whereas in typical experiments $k_0, q_0 \gtrsim$ MeV in particular, for the decay of mesons, if the localization length scale is of the order of the decay length the typical ratio $k_0/\sigma, q_0/\sigma \gtrsim 10^{12}$ and typical energy (momentum) resolutions are $\gg \sigma$. Therefore for all experimental intent and purpose the wave packets are orthogonal for different values of the momentum

$$\langle \alpha; \vec{q}_0; \vec{x}_0 | \alpha; \vec{k}_0; \vec{x}_0 \rangle \simeq \delta_{\vec{q}_0, \vec{k}_0} \, . \tag{4.4.13}$$

From the identity (4.4.8) an important property of these wave packets that will be useful below is the following identity,

$$\sum_{\vec{k}} C_\alpha(\vec{k} - \vec{k}_0; \vec{x}_0) \, |1^\alpha_{\vec{k}-\vec{q}}\rangle = |\alpha; \vec{k}_0 - \vec{q}; \vec{x}_0\rangle \, . \tag{4.4.14}$$

Although this property is evident with the Gaussian wave packets (4.4.9) it is quite general for localized functions of $\vec{k} - \vec{k}_0$.

4.4.2 Time evolution:

Consider the single particle wavepacket (4.4.2) with $C_\alpha(\vec{k};\vec{k}_0;\vec{x}_0)$ given by (4.4.9). The time evolution of this state is given by

$$|\alpha;\vec{k}_0;\vec{x}_0;t\rangle = e^{-iH_0t}|\alpha;\vec{k}_0;\vec{x}_0\rangle = \sum_{\vec{k}} C_\alpha(\vec{k}-\vec{k}_0;\vec{x}_0)\, e^{-iE_\alpha(k)t}|1^\alpha_{\vec{k}}\rangle\,. \tag{4.4.15}$$

For a wavepacket sharply localized in momentum, we can expand

$$E_\alpha(k) = E_\alpha(k_0) + \vec{V}_g(k_0)\cdot(\vec{k}-\vec{k}_0) + \cdots\,, \tag{4.4.16}$$

where

$$\vec{V}_g(k_0) = \frac{\vec{k}_0}{E_\alpha(k_0)} \tag{4.4.17}$$

is the group velocity, the second derivative terms in (4.4.16) give rise to transverse and longitudinal dispersion. Neglecting both transverse and longitudinal dispersion under the assumption that the packet is narrowly localized in momentum and the time scales are much shorter than the dispersion scales (see discussion below), it is straightforward to find

$$|\alpha;\vec{k}_0;\vec{x}_0;t\rangle = e^{-iE_\alpha(k_0)t}\,|\alpha;\vec{k}_0;(\vec{x}_0-\vec{V}_g(k_0)t)\rangle\,, \tag{4.4.18}$$

namely, neglecting dispersion the center of the wave packet moves with the group velocity as expected.

4.4.3 N-particle wavepackets:

We have normalized the wavepackets to describe one particle in the volume V as is evident from the result (4.4.4). However, for experimental purposes one may consider initial states with a single particle within the *localization volume* $1/\sigma^3$. Because the quantum field theory is quantized in a (much larger) volume V, these initial states must, therefore, feature different normalization. This can be seen from the normalization of the single particle states (4.4.3) with the usual passage to the continous momentum description

$$\sum_{\vec{k}} \to V \int \frac{d^3k}{(2\pi^3)}$$

which explains the volume factor in (4.4.9). The condition of one single particle state per localization volume therefore requires a different normalization of the single particle wavepackets, which can be obtained by dividing up the total volume V into bins of volume $1/\sigma^3$ with one single particle in each bin, leading to $N = V\sigma^3$ total particles in the volume V, with a particle density in the volume V given by $N/V = 1/(1/\sigma^3) = \sigma^3$, so that the total normalization in the whole volume V must be N.

Therefore, we can accomplish the description of the initial state of the decaying parent particle in terms of a wavepacket of single particle states with one single particle within the localization volume by normalizing these states to N in the volume V. Namely we consider the initial wavepacket states in terms of the momentum wavefunctions

$$C_{N\alpha}(\vec{k} - \vec{k}_0; \vec{x}_0) = \sqrt{N}\, C_\alpha(\vec{k} - \vec{k}_0; \vec{x}_0)\,, \tag{4.4.19}$$

in which case the Gaussian wavepacket (4.4.9) becomes

$$C_{N\alpha}(\vec{k} - \vec{k}_0; \vec{x}_0) = \left[\frac{8\,\pi^{\frac{3}{2}}\,n}{\sigma^3}\right]^{\frac{1}{2}} e^{-\frac{(\vec{k}-\vec{k}_0)^2}{2\sigma^2}}\; e^{i(\vec{k}-\vec{k}_0)\cdot\vec{x}_0}\,, \tag{4.4.20}$$

where $n = N/V = \sigma^3$ is the particle *density*. The corresponding quantum states

$$|N\alpha; \vec{k}_0, \vec{x}_0\rangle = \sum_{\vec{k}} C_{N\alpha}(\vec{k} - \vec{k}_0; \vec{x}_0)\, |1^\alpha_{\vec{k}}\rangle = \sqrt{N}|\alpha; \vec{k}_0, \vec{x}_0\rangle\,. \tag{4.4.21}$$

Now the localized states (4.4.2) are normalized to the total number of particles in the volume V, namely $N = V\sigma^3$, (with one particle per localization volume) instead of (4.4.3), and they are still orthogonal in the sense of (4.4.13).

We will consider the entanglement entropy within a localization volume rather than in the total volume.

4.4.4 Wigner-Weisskopf with wave packets:

This wave packet description is easily incorporated into the Wigner-Weisskopf approach to the description of the full time evolution of the quantum state of the decaying parent particle, the quantum state in the interaction picture (4.2.4) is generally written as

$$|\Psi(t)\rangle = \sum_{\vec{k}} C_\Phi(\vec{k}, \vec{k}_0; \vec{x}_0; t)|1^\Phi_{\vec{k}}\rangle + \sum_\kappa \mathcal{C}_\kappa(t)|\kappa\rangle \tag{4.4.22}$$

where the states $|\kappa\rangle$ are multiparticle states, with the initial conditions

$$C_\Phi(\vec{k}; \vec{k}_0; \vec{x}_0; t = 0) = C_\Phi(\vec{k} - \vec{k}_0; \vec{x}_0) \quad ; \quad \mathcal{C}_\kappa(t = 0) = 0\,, \tag{4.4.23}$$

where $C_\Phi(\vec{k} - \vec{k}_0; \vec{x}_0)$ describe the localized wave packet of the decaying parent particle at the initial time. The interaction Hamiltonian connects the single particle plane wave states $|1^\Phi_{\vec{k}}\rangle$ with the two-particle plane wave states $|1^\chi_{\vec{p}}\rangle\,|1^\psi_{\vec{k}-\vec{p}}\rangle$ with matrix element

$$\mathcal{M}_\Phi(\vec{k}, \vec{p}) = \langle 1^\Phi_{\vec{k}}|\hat{H}_I(0)|\chi_{\vec{p}}\,\psi_{\vec{k}-\vec{p}}\rangle = \frac{g}{\sqrt{8VE_\Phi(k)E_\psi(|\vec{k} - \vec{p}|)E_\chi(p)}} \tag{4.4.24}$$

leading to the decay rate

$$\Gamma_k = \frac{\Gamma m_\Phi}{E_\Phi(k)} \tag{4.4.25}$$

with Γ is the decay rate in the rest frame of the parent particle given by eqn. (4.3.14). Following the same steps as described in the previous section we now find

$$C_\Phi(\vec{k}, \vec{k}_0; \vec{x}_0; t) = C_\Phi(\vec{k} - \vec{k}_0; \vec{x}_0)\, e^{-i\Delta E_\Phi(k)\, t}\, e^{-\Gamma_k t/2} \tag{4.4.26}$$

where $\Delta E_\Phi(k) = \delta m^2_\Phi/2E_\Phi(k)$ and

$$\mathcal{C}_{\chi\psi}(\vec{k};\vec{p};t) = C_\Phi(\vec{k}-\vec{k}_0;\vec{x}_0)\,\mathcal{M}_\Phi(\vec{k},\vec{p})\,\frac{\Big[1-e^{-i(E_\Phi(k)-E_\chi(p)-E_\psi(|\vec{k}-\vec{p}|)-i\Gamma_k/2)\,t}\Big]}{\Big[E_\psi(p)+E_\psi(|\vec{k}-\vec{p}|)-E_\Phi(k)+i\Gamma_k/2\Big]}\,. \quad (4.4.27)$$

and to leading order in the interaction we find

$$|\Psi(t)\rangle = \sum_{\vec{k}} C_\Phi(\vec{k},\vec{k}_0;\vec{x}_0;t)|1^\Phi_{\vec{k}};0_\chi;0_\psi\rangle + \sum_{\vec{k},\vec{p}} \mathcal{C}_{\chi\psi}(\vec{k};\vec{p};t)|1^\chi_{\vec{p}};1^\psi_{\vec{k}-\vec{p}};0^\Phi\rangle\,. \quad (4.4.28)$$

The number of $\Phi;\psi$ particles respectively are given by

$$N^\Phi(t) = \langle\Psi(t)|\sum_{\vec{q}} a^\dagger_{\Phi,\vec{q}}\,a_{\Phi,\vec{q}}|\Psi(t)\rangle = \sum_{\vec{k}}|C_\Phi(\vec{k},\vec{k}_0;\vec{x}_0;t)|^2 \quad (4.4.29)$$

$$N^\psi(t) = \langle\Psi(t)|\sum_{\vec{q}} a^\dagger_{\psi,\vec{q}}\,a_{\psi,\vec{q}}|\Psi(t)\rangle = \sum_{\vec{p}}\sum_{\vec{k}}|C_{\chi,\psi}(\vec{k},\vec{p};t)|^2\,. \quad (4.4.30)$$

In order to understand the physical consequences of the wave packet description in the clearest manner, let us consider Gaussian wave packets localized at the origin with vanishing average momentum, namely $\vec{k}_0 = 0;\vec{x}_0 = 0$, and describing a single particle in the volume V so as to establish contact with the plane wave results from the previous section, namely

$$C_\Phi(\vec{k};\vec{0};\vec{0}) = \left[\frac{8\,\pi^{\frac{3}{2}}}{\sigma^3\,V}\right]^{\frac{1}{2}}\,e^{-\frac{k^2}{2\sigma^2}}\,. \quad (4.4.31)$$

The main assumption in what follows is that the spatial localization length $1/\sigma$ be much larger than the Compton wavelength of the decaying particle $1/m_\Phi$, namely our *main assumption* on the property of the wave packets is that

$$\frac{\sigma}{m_\Phi} \ll 1\,. \quad (4.4.32)$$

In the case of meson decay this is physically correct as any localization length smaller than the Compton wavelength will necessarily explore the inner structure of the decaying particle and would be sensitive to the short distance compositeness scale. However as we argue below, this assumption is more general, since a wave packet of massive particles localized on distances shorter than the Compton wavelength will disperse on time scales shorter than the particle's oscillation scale.

Consider the first term in (4.4.22), it will contribute density matrix elements that will feature typical integrals of the form

$$I = \left[\frac{8\,\pi^{\frac{3}{2}}}{\sigma^3\,V}\right]^{\frac{1}{2}} \int \frac{d^3k}{(2\pi)^3}\, e^{-k^2/2\sigma^2} e^{-i\Delta E_\Phi(k)t} e^{-\Gamma_k t/2} \tag{4.4.33}$$

changing variables to $s = k/\sigma$ it follows that

$$\Delta E_\Phi(k) = \Delta E_\Phi(0)\Big[1 - \frac{s^2}{2}\,\frac{\sigma^2}{m_\Phi^2} + \cdots\Big] \;\; ; \;\; \Gamma_k = \Gamma\Big[1 - \frac{s^2}{2}\,\frac{\sigma^2}{m_\Phi^2} + \cdots\Big]\,, \tag{4.4.34}$$

where Γ is the decay rate of the particle at rest. The integrand is strongly suppressed for $|s| > 1$ and for $\sigma^2/m_\Phi^2 \ll 1$ the corrections inside the brackets can be systematically computed in a power series expansion, yielding corrections of the form

$$\Gamma t\Big[\sigma^2/m_\Phi^2 + \cdots\Big] \;\; ; \;\; -imt\Big[\sigma^2/m_\Phi^2 + \cdots\Big]\,. \tag{4.4.35}$$

The corrections to the decay are negligible during the lifetime of the decaying particle $\Gamma t \simeq 1$ and can be safely neglected. To lowest order in σ^2/m_Φ^2 the corrections to the energy can be absorbed into a time dependent width

$$\sigma^2(t) \simeq \frac{\sigma^2}{\big[1 + i\frac{\sigma^2}{m_\Phi^2}(m_\Phi t)\big]} \tag{4.4.36}$$

describing the dispersion of the wave packet. For $\sigma^2/m_\Phi^2 \ll 1$ we can also neglect the dispersion of the wave packet over the time scale of many oscillations, with the result that

$$I \propto e^{-i\Delta E_\Phi(0)t} e^{-\Gamma t/2}\Big[1 + \mathcal{O}(\sigma^2/m_\Phi^2)\Big]\,, \tag{4.4.37}$$

this is the justification for neglecting the dispersion in the time-evolved wave packet assumed in section (4.4.2).

Therefore, when considered under integrals (or discrete sums) we can safely replace

$$C_\Phi(\vec{k};\vec{0};\vec{0};t) \to e^{-i\Delta E_\Phi(0)t} e^{-\Gamma t/2}\, C_\Phi(\vec{k};\vec{0};\vec{0};0)\,. \tag{4.4.38}$$

Physically the result above is the statement that for $\sigma^2/m_\Phi^2 \ll 1$ the dispersion can be safely neglected during the lifetime of the decaying state. For example, consider instead that $\sigma/m_\Phi \simeq 1$, then the wavepacket will disperse within a time scale given by $m_\Phi t \simeq m_\Phi^2/\sigma^2 \simeq 1$,

namely the wavepacket would completely disperse within one oscillation and the concept of the decay rate is not relevant as the amplitude of the wave packet diminishes quickly by dispersion and not by decay.

In particular to leading order in σ^2/m_Φ^2 we find

$$N^\Phi(t) = e^{-\Gamma t} \quad ; \quad N^\psi(t) = 1 - e^{-\Gamma t}\,, \tag{4.4.39}$$

which are the same results as for the plane wave case. Implementing these approximations, to leading order in this ratio, the first term in (4.4.22) becomes

$$\begin{aligned}\sum_{\vec{k}} C_\Phi(\vec{k},\vec{0};\vec{0};t)|1^\Phi_{\vec{k}}\rangle &= e^{-i\Delta E_\Phi(0)t}\, e^{-\Gamma t/2} \sum_{\vec{k}} C_\Phi(\vec{k},\vec{0};\vec{0};t=0)|1^\Phi_{\vec{k}}\rangle + \mathcal{O}(\sigma^2/m_\Phi^2) \\ &= e^{-i\Delta E_\Phi(0)t}\, e^{-\Gamma t/2}|\Phi;\vec{0};\vec{0}\rangle + \mathcal{O}(\sigma^2/m_\Phi^2)\end{aligned} \tag{4.4.40}$$

where $|\Phi;\vec{0};\vec{0}\rangle$ is a zero momentum wave packet state localized at the origin.

Therefore we clearly see that when $\sigma^2/m_\Phi^2 \ll 1$, namely when the localization length scale is much larger than the Compton wavelength of the particle we obtain the same result as in the plane wave case but with the replacement

$$|1^\Phi_{\vec{k}}\rangle \to |\Phi;\vec{k};\vec{x}_0\rangle\,. \tag{4.4.41}$$

The same argument is applied to integrals of the form

$$\int d^3k\, \mathcal{C}_{\chi\psi}(\vec{k};\vec{p};t)$$

with $\mathcal{C}_{\chi\psi}(\vec{k};\vec{p};t)$ given by (4.4.27), writing $k = s\sigma$ with the integration range $|s| \lesssim 1$ the k-dependent terms can be expanded around $k = 0$, and the k-dependent terms yield corrections in powers of $\sigma/m_\Phi \ll 1$, the leading order is given by the $k = 0$ contribution, which is obtained by the simple replacement

$$\mathcal{C}_{\chi\psi}(\vec{k};\vec{p};t) \to C_\Phi(\vec{k};\vec{0};\vec{0};t=0)\,\mathcal{C}_{\chi\psi}(p;t) \tag{4.4.42}$$

where $\mathcal{C}_{\chi\psi}(p;t)$ is the *plane wave result* (4.3.17).

The reduced density matrix is obtained by tracing $\rho(t)$ given by (4.3.21) over the χ degrees of freedom, namely

$$\begin{aligned}\rho_\psi(t) = \mathrm{Tr}_\chi \rho(t) &= \sum_{\vec{k},\vec{k}'} C^{\Phi}(\vec{k};\vec{0},\vec{0},t)\, C^{*\,\Phi}(\vec{k}';\vec{0},\vec{0},t)|1^{\Phi}_{\vec{k}}\rangle\langle 1^{\Phi}_{\vec{k}'}| \\ &+ \sum_{\vec{k},\vec{k}',\vec{p}} \mathcal{C}_{\chi\psi}(\vec{k};\vec{p};t)\, \mathcal{C}^*_{\chi\psi}(\vec{k}';\vec{p};t)|1^{\psi}_{\vec{k}-\vec{p}}\rangle\langle 1^{\psi}_{\vec{k}'-\vec{p}}|\,, \end{aligned} \tag{4.4.43}$$

to leading order in σ^2/m^2_Φ and using the results (4.4.38,4.4.42) we find

$$\begin{aligned}\rho_\psi(t) &= e^{-\Gamma t}\Big(\sum_{\vec{k}} C^{\Phi}(\vec{k};\vec{0},\vec{0},0)|1^{\Phi}_{\vec{k}}\rangle\Big)\Big(\sum_{\vec{k}'} C^{*\,\Phi}(\vec{k}';\vec{0},\vec{0},0)\langle 1^{\Phi}_{\vec{k}'}|\Big) \\ &+ \sum_{\vec{p}} |\mathcal{C}_{\chi\psi}(p;t)|^2 \Big(\sum_{\vec{k}} C^{\Phi}(\vec{k};\vec{0},\vec{0},0)|1^{\psi}_{\vec{k}-\vec{p}}\rangle\Big)\Big(\sum_{\vec{k}'} C^{*\Phi}(\vec{k};\vec{0},\vec{0},0)\langle 1^{\psi}_{\vec{k}'-\vec{p}}|\Big) \end{aligned} \tag{4.4.44}$$

however

$$\sum_{\vec{k}} C^{\Phi}(\vec{k};\vec{0},\vec{0},0)|1^{\Phi}_{\vec{k}}\rangle = |\Phi;\vec{0};\vec{0}\rangle \tag{4.4.45}$$

is the original wave packet of the parent particle with zero average momentum and localized at the origin describing a wave packet for the parent particle at rest, and using the property (4.4.14) we find

$$\sum_{\vec{k}} C^{\Phi}(\vec{k};\vec{0},\vec{0},0)|1^{\psi}_{\vec{k}-\vec{p}}\rangle = |\psi;-\vec{p};\vec{0}\rangle\,, \tag{4.4.46}$$

this is a wave packet of daughter particles with average momentum $-\vec{p}$ localized at the origin with spatial localization length $1/\sigma$ with the same wave packet profile as the parent particle. Namely, the daughter particles "inherit" the wave packet structure of the parent particle. Therefore to leading order in σ^2/m^2_Φ we finally find (after relabelling $-\vec{p}\to\vec{p}$)

$$\rho_\psi(t) = e^{-\Gamma t}|\Phi;\vec{0};\vec{0}\rangle\langle\Phi;\vec{0};\vec{0}| + \sum_{\vec{p}} |\mathcal{C}_{\chi\psi}(p;t)|^2\, |\psi;\vec{p};\vec{0}\rangle\langle\psi;\vec{p};\vec{0}| \tag{4.4.47}$$

Remarkably, *this is the same result as in the plane wave case* eqn. (4.3.27) but with the replacement of the single particle plane wave states by the respective localized wavepackets with the corresponding (average) momenta. The corrections are of $\mathcal{O}(\sigma^2/m^2_\Phi)$ as discussed above.

The density matrix (4.4.47) is in the *interaction picture* where the only time dependence is from the interaction and encoded in the decaying exponential and the Wigner-Weisskopf coefficients. The density matrix in the *Schroedinger picture* is given by

$$\rho_{\psi}^{(S)}(t) = e^{-iH_0 t}\ \rho_{\psi}(t)\ e^{iH_0 t}\,. \tag{4.4.48}$$

The application of the free time evolution operator on the wave packets yields the wavepackets with the centers displaced by $-\vec{V}_g t$ where the group velocity vanishes for the wavepacket of the decaying particle but it is $\vec{V}_g(p)$ for the ψ wavepackets. Thus the center of the wavepackets moves with the group velocity (again neglecting dispersion).

However, our goal is to obtain the entanglement entropy, for which the unitary transformation (4.4.48) is irrelevant. Hence insofar as the entanglement entropy is concerned, the motion of the center of the wave packet does not affect the result.

Although the density matrix seems diagonal in the wavepacket description, the wavepackets are not exactly orthogonal for different values of the momenta, so the density matrix in the wave packet representation in principle features off-diagonal matrix elements. However, the coefficient $|\mathcal{C}_{\chi\psi}(p;t)|^2$ is strongly peaked at a value of the momentum p^* which determined by the kinematics of the decay into plane wave states (the eigenfunctions of the Hamiltonian) from the parent particle at rest, with a Lorentzian profile whose width is the lifetime of the parent particle Γ. Therefore only states with $p \simeq p^*$ within a region of width Γ contribute, typically p^* few MeV. However, from the results of section (4.4.1) above, wavepackets that are localized on a macroscopic distance are nearly orthogonal in the sense of eqn. (4.4.13). Therefore the wave packets furnish an (nearly) orthonormal set and the density matrix is (nearly) diagonal in these states, off diagonal elements only contribute within a width σ and their contributions are suppressed by powers of σ/p^*.

The calculation of the entanglement entropy now proceeds just as in the previous section with the final result given by eqn. (4.3.30).

The fact that the density matrix for the wave-packet description is similar to that of the plane wave description in the regime where the spatial localization scale of the wave packet is much larger than the Compton wavelength of the particle is expected on physical grounds.

For example in formal S-matrix theory, the correct approach to describing a scattering event is in terms of localized wave packets for the projectile and target particles prepared in the far past and evolved into the far future. However an actual calculation of a scattering cross section is performed in terms of plane waves, with equal probability everywhere in space. Furthermore, the asymptotic reduction formula that allows to extract S-matrix elements from Green's functions invokes asymptotic in and out states in terms of the single particle eigenstates of the unperturbed Hamiltonian of definite energy and momentum, these being the states that transform as irreducible representations of the Lorentz group. Wave packets do *not* feature definite energy and momenta, yet they underlie all formal descriptions of scattering theory. These two approaches are reconciled when the wave packets of the incoming and outgoing particles are localized on distances larger than the Compton (or de Broglie, whichever is shorter) wavelengths. Similarly, the decay rate of a particle is typically calculated in the plane wave basis, and a particle decaying at rest is assigned a decay rate at zero momentum. However in an experimental situation decaying states are produced generally as wavepackets which are superpositions of plane wave states, each of which would decay with a different rate, yet the description of a decaying particle at rest is in terms of one decay rate extracted as the transition probability per unit time. The decay of a wave packet would entail several different decay time scales. Again this situation is reconciled by considering localized wavepackets but whose localization length is much larger than the typical scale of the particle, namely the Compton or de Broglie wavelength whichever is shorter.

4.4.5 Wave packets of finite particle density:

In the treatment of the previous section, we have considered wave packets that describe a single particle in the volume V and explained that this is the reason that the entanglement entropy depends logarithmically on the volume. However, experimentally and more physically we should describe a state that has one particle *in the localization volume* and not in the total volume. This is achieved by considering the wavepackets describing N-single particles described in section (4.4.3), namely the states (4.4.21). The Wigner-Weisskopf method follows exactly the same steps as before, leading to density matrix in the interaction picture

of the same form as (4.4.47) but with the wavepackets $|N\Phi;\vec{0};\vec{0}\rangle = \sqrt{N}|\Phi;\vec{0};\vec{0}\rangle$; $|N\psi;\vec{p};\vec{0}\rangle = \sqrt{N}|\psi;\vec{p};\vec{0}\rangle$, therefore $\rho(t) \to N\rho(t)$ leading to the following entanglement entropy

$$S_N(t) = -N\left[e^{-\Gamma t}\ln\left[N\,e^{-\Gamma t}\right] + \left[1-e^{-\Gamma t}\right]\ln\left[1-e^{-\Gamma t}\right] + \left[1-e^{-\Gamma t}\right]\ln\left[Nn^{\psi}(p^*\,;\infty)\right]\right] \quad (4.4.49)$$

where

$$Nn^{\psi}(p^*\,;\infty) = \frac{4\pi\, m_\Phi\, n}{p^* E_\chi(p^*) E_\psi(p^*)\Gamma}\,, \quad (4.4.50)$$

where $n = N/V$ is the particle density.

Assuming that there is one particle *per localization volume* $1/\sigma^3$, it follows that $n = \sigma^3$ and asymptotically the entanglement entropy is given by

$$S_N(\infty) = -N\ln\left[\frac{4\pi\, m_\Phi\, \sigma^3}{p^* E_\chi(p^*) E_\psi(p^*)\Gamma}\right]. \quad (4.4.51)$$

Thus we see, that as anticipated by the discussion above, the entanglement entropy is extensive, the volume dependence has now been replaced by the localization volume, in particular the specific entanglement entropy (entropy per unit total volume) is

$$s_N(\infty) = \frac{S_N(\infty)}{V} = -\sigma^3\ln\left[\frac{4\pi\, m_\Phi\, \sigma^3}{p^* E_\chi(p^*) E_\psi(p^*)\Gamma}\right], \quad (4.4.52)$$

and within a localization volume $1/\sigma^3$ we find

$$S_{loc}(\infty) = -\ln\left[\frac{4\pi\, m_\Phi\, \sigma^3}{p^* E_\chi(p^*) E_\psi(p^*)\Gamma}\right]. \quad (4.4.53)$$

This is the *same* result as in the plane wave case (4.3.32) but with the localization volume replacing the total volume, which is physically reasonable, for an arbitrary *density* of particles $n = N/V$ the factor σ^3 is replaced by n and for one particle in the volume V one recovers the result (4.3.32).

4.5 INTERPRETATION AND A *POSSIBLE* EXPERIMENTAL MEASUREMENT

The logarithmic dependence of the entanglement entropy (4.3.30) on the volume factor has a clear statistical interpretation. Consider a dilute gas of particles whose statistical distribution or phase space density is f_p. The total *density* of particles is

$$\frac{N}{V} = \int \frac{d^3k}{(2\pi)^3}\, f_p \tag{4.5.1}$$

and the Von-Neumann entropy of this (dilute) gas is

$$S_{VN} = -\sum_p f_p \ln[f_p] = -V \int \frac{d^3k}{(2\pi)^3}\, f_p \, \ln[f_p]\,. \tag{4.5.2}$$

If the number of particles remains finite in the large volume limit, namely if the particle density scales $\propto 1/V$ in this limit, then it follows that $f_p \propto 1/V$. On the contrary, if f_p is independent of the volume as in the cases of the Maxwell-Boltzmann, Bose-Einstein or Fermi-Dirac distributions, the total density is *finite* in the infinite volume limit and the entropy is extensive. For a finite number of particles (vanishing particle density in the infinite volume limit) $f_p \propto 1/V$ and the Von-Neumann entropy is *not* extensive,

$$S_{VN} \propto N \ln[V]\,. \tag{4.5.3}$$

This is *precisely* the origin of the logarithmic dependence on the volume of the entanglement entropy (4.3.32) for the case of a single particle in the volume V: the initial state has one particle and the final state has one (of each) daughter particle, the distribution function of the daughter particles at asymptotically long times after the decay of the parent particle is $|\mathcal{C}_{\chi\psi}(p,\infty)|^2 \propto 1/V$ the inverse volume dependence is the statement that there is a finite number of particles distributed in phase space. Obviously this volume dependence is independent of whether the states are described by plane waves or wave packets, but is a statement of the simple fact that the number of particles in the volume V is finite.

When we consider wavepackets describing a fixed number of particles *per localization volume*, this case corresponds to a finite density, and for one particle per localization volume

it follows that the entanglement entropy becomes extensive and the volume dependence in the logarithm is replaced by the localization volume $1/\sigma^3$,

$$S_{VN} \propto N \ln[\sigma^3] \,. \tag{4.5.4}$$

The entanglement entropy from the decay of a parent particle is *in principle* experimentally accessible: consider the typical experiment in which a pion is produced at rest from protons incident on a target and the entangled muon-neutrino pairs from pion decay are distributed isotropically. Consider that muons are detected with a 4π detector within the pion's decay region but not the neutrinos. By counting the number of muons within momenta bins of resolution Δp the phase space density, namely the muon statistical distribution function is the number of muons detected within this momentum "bin" per unit physical detection volume. This is the quantity f_p, the entanglement entropy is the Von-Neumann statistical entropy $S = -\sum_{all\ bins} f_p \ln[f_p]$ and is a measure of the *information loss* resulting from tracing out the neutrino degrees of freedom and detecting only muons. As an order of magnitude estimate consider localization within a detection volume of the order of $(c\tau)^3$ namely with $\sigma = \Gamma$ in eqn. (4.4.53) with $c\tau \approx 7.8\,\text{mts}$ being the decay length of the pion at rest, the muon and neutrino are emitted with a Lorentzian distribution of energy peaked at momenta $p^* = 30\,\text{MeV}$ and $m_\nu \ll m_\mu$ and width Γ the pion's decay rate, we find the asymptotic entanglement entropy within this volume to be $S_{loc}(\infty) \simeq 70$.

4.6 CONCLUSIONS AND FURTHER QUESTIONS

The decay of a parent particle leads to a quantum entangled state of the daughter particles as a consequence of conservations laws in the decay process. Experiments at Belle[156] measured (EPR) correlations in entangled pairs of B mesons and further experiments at Belle and Babar[157, 158] exploit the correlations in entangled B-meson pairs to study CP and T violation by tagging members of the pairs and studying the time evolution of flavor asymmetries. Further proposals suggest to extend these studies exploiting entanglement and correlations to measure CP and T violating observables in other B-meson systems.

Motivated by these timely experiments that access quantum correlations in entangled states to extract physical information, we focus on a *complementary* aspect of quantum entanglement of particles produced from the decay of a parent particle: if one of the members of the correlated state cannot or will not be measured, tracing over its dynamical degrees of freedom results in a *reduced* density matrix. The Von-Neumann entropy associated with this mixed density matrix, namely the entanglement entropy, measures the loss of information that was originally contained in the (EPR) correlations of the entangled state.

We generalized and extended a method used in the study of spontaneous decay of atomic systems to the realm of quantum field theory to obtain in a consistent approximation, the full quantum state that describes the time evolution of the decaying particle and the production of the daughter particles. This method is non-perturbative and is manifestly unitary. We have implemented the method to study the simpler case of bosonic parent and daughter particles to highlight the main concepts and consequences, however, the results are quite general.

The full quantum state resulting from the time evolution of the decaying particle yields a pure state density matrix. However, if one (or more) daughter particles is unobserved, tracing over their degrees of freedom leads to a mixed state density matrix whose time evolution is completely determined by the *unitary* time evolution of the decay process. This mixed state density matrix features an *entanglement entropy* which is a manifestation of the quantum correlations of the entangled product state. We obtained the time evolution of the entanglement entropy and show that it grows on a time scale determined by the lifetime of the decaying particle and reaches a maximum that corresponds to the logarithm of the available phase space states of the decay particles. For a parent particle described by a narrow resonance the distribution of produced (entangled) daughter particles is nearly a constant in a narrow energy-momentum region, the emitted particles are nearly maximally entangled Bell-states.

We have extended the study to the case in which the decaying parent particle is described by a wave packet, rather than a plane wave. The daughter particles "inherit" the wave packet structure of parent particle. We have demonstrated the equivalence between the reduced

density matrix in terms of plane waves and that in terms of localized wavepackets under the physically reasonable approximation that the localization length $1/\sigma$ of the decaying state is much larger than its Compton wavelength $1/m$ with corrections of $\mathcal{O}(\sigma^2/m^2) \ll 1$.

Furthermore, we have discussed *possible* experimental ways to access the entanglement entropy. Although unobserved states are manifest as missing energy, the entanglement entropy provides a complementary tool that measures the loss of information contained in the original quantum correlations between the members of the pair of particles produced in the decay process. Just as the measurement of (EPR) correlations at Belle[156] provide a confirmation of fundamental aspects of quantum mechanics, now accessed at B-factories, a measurement or confirmation of the entanglement entropy could provide yet another complementary test.

While experiments in quantum optics are testing fundamental concepts associated with the entanglement entropy and applying them for quantum information platforms[151, 149, 150], these concepts have not *yet* received the attention of the particle physics community but it is conceivable that they *may* prove relevant in the statistical analysis of the time evolution of entangled B-meson pairs in studies of CP violation.

This work is also a prelude to the assessment of the entanglement entropy due to particle decays in de Sitter space[175]. There, due to the fact that particle can decay into itself with momenta that are much less than the Hubble constant $H_{\rm deSitter}$[175], we expect there to be an interplay between the horizon size and the decay rate that will feed in to the behavior of the entanglement entropy. This may then mix the ideas of entanglement entropy developed here with those coming from studies of spatially separated portions of de Sitter space such as in ref.[172]. Work on these aspects will be reported elsewhere[198].

5.0 SUPERHORIZON ENTANGLEMENT ENTROPY FROM PARTICLE DECAY IN INFLATION.

Based on: (ref. [299])

L. Lello, D. Boyanovsky, R. Holman, JHEP 04 055 (2014)

5.1 INTRODUCTION

Quantum fluctuations during inflation seed the anisotropies in the cosmic microwave background and generate primordial gravitational waves. In its simplest inception the inflationary stage can be effectively described as a quasi-deSitter space time. Early studies[177, 178, 179, 180, 181, 182] revealed that de Sitter space time features infrared instabilities and profuse particle production in interacting field theories. During inflation the rapid cosmological expansion modifies the energy-uncertainty relation allowing "virtual" excitations to persist longer, leading to remarkable phenomena, which is stronger in de Sitter space time[183]. Particle production in a de Sitter background has been argued to provide a dynamical"screening" mechanism that leads to relaxation of the cosmological constant[184, 185, 186] through back reaction, much like the production of particle-antiparticle pairs in a constant electric field. More recently this mechanism of profuse particle production has been argued to lead to the instability of de Sitter space time[187, 188].

A particular aspect of the rapid cosmological expansion is the lack of a global time-like killing vector which leads to remarkable physical effects in de Sitter space time, as it implies

the lack of kinematic thresholds (a direct consequence of energy-momentum conservation) and the decay of fields even in their own quanta[229, 190, 175] with the concomitant particle production. This result was confirmed in ref.[191, 192, 193] and more recently in ref.[194] by a thorough analysis of the S-matrix in global de Sitter space time.

The decay of an initial single particle state into many particle states results in a quantum state that is kinematically entangled in momentum space: consider the example of a scalar field theory with cubic self-interaction and an initial single particle state with spatial physical momentum $\vec{k}$, namely $|1_{\vec{k}}\rangle$, this state decays into a two-particle states of the form $\sum_{\vec{p}} C_{\vec{p}}(t)\,|1_{\vec{p}}\rangle|1_{\vec{k}-\vec{p}}\rangle$ where $C_{\vec{p}}(t)$ is the *time dependent* amplitude of the two particle state with momenta $\vec{p}$ and $\vec{k}-\vec{p}$ respectively. This is an entangled state that features non-trivial correlations between the product particles. In ref.[229, 175] it is argued that in de Sitter space time with Hubble constant H, the largest decay amplitude corresponds to the case when one of the product particles features physical momenta $p \ll H$, therefore, if the initial particle has physical momenta $k \gg H$ and one of the product particles features a momentum $p \ll H$ (the other with $|\vec{k}-\vec{p}| \gg H$) the quantum entangled state features correlations between the sub and superHubble daughter particles.

We refer to these correlated pairs produced from the decay of a parent particle as entangled across the Hubble radius, namely "superhorizon" entanglement, referring to the Hubble radius in de Sitter space time as the horizon as is customary in inflationary cosmology.

Correlations of quantum fluctuations during a de Sitter inflationary stage have been recently argued[195] to lead to remarkable Hanbury-Brown-Twiss interference phenomena with potential observational consequences.

Unitary time evolution of an initial single particle state is a pure quantum state in which the product particles are kinematically entangled. If a pure quantum state describes an entangled state of several subsystems and if the degrees of freedom of one of the subsystems are not observed, tracing the pure state density matrix over these unobserved degrees of freedom leads to a *mixed state reduced density matrix*. The entanglement entropy is the Von Neumann entropy associated with this reduced density matrix; it reflects the loss of information that was originally present in the quantum correlations of the entangled state.

The main purpose of this article is to study the entanglement entropy in the case of an initial quantum state describing a single particle state with physical momentum $k \gg H$ decaying into a pair of particles one with $p \ll H$ (superhorizon), and the other with $|\vec{k} - \vec{p}| \gg H$ (subhorizon) by tracing over the super-Hubble ("superhorizon") degrees of freedom. This entanglement entropy is a measure of the loss of information contained in the pair correlations of the daughter particles.

The entanglement entropy has been the focus of several studies in condensed matter systems[154, 155, 196, 164], statistical physics and quantum field theory[163, 167, 165, 166, 168], black hole physics[169, 170, 197] and in particle production in time dependent backgrounds[171]. Most of these studies focus on entanglement between spatially correlated regions across boundaries. The entanglement entropy in de Sitter space-time for a free, minimally coupled massive scalar field has been studied in ref.[172] with the goal of understanding superhorizon correlations, and ref.[173] studied the entropy from momentum space entanglement and renormalization in an interacting quantum field theory in Minkowski space-time.

Our study differs from these studies in many ways: we are not considering spatially correlated regions, and momentum space entanglement resulting from the kinematics of particle decay in states of the same quanta is different from the cases studied in ref.[173] which considered momentum space entanglement in the interacting *ground* state of a quantum field theory or a finite density case, both in a stationary, equilibrium situation, whereas we are interested in the time evolution of the reduced density matrix and the concomitant increase of the entanglement entropy in an interacting theory in de Sitter space time.

More recently the entanglement entropy in the ubiquitous case of particle decay in Minkowski space-time from tracing over the degrees of freedom of an unobserved daughter particle has been studied in ref.[198] as a characterization of an "invisible" decay complementary to missing energy.

We focus on light scalar fields with mass $M^2 \ll H^2$, for which radiative corrections feature infrared divergences that are manifested as poles in $\Delta = M^2/3H^2 \ll 1$[229, 175] in the self-energy leading to a consistent expansion in Δ. A similar expansion was recognized

in refs.[199, 200, 201, 202].

The field theoretic method introduced in ref.[175, 198] that describes the *non-perturbative* time evolution of quantum states is extended here and then generalized to inflationary cosmology (for other applications of this field theoretical method see refs.[162]) to obtain the entangled quantum state from single particle decay to leading order in a $\Delta = M^2/3H^2$ expansion. We show explicitly that unitarity is manifest in the time evolution of the quantum state. From this state we construct the (pure) density matrix and trace over the contribution from superhorizon modes and obtain the entanglement entropy to leading order in a Δ expansion. Whereas in ref.[203] the entanglement between *only two modes* was studied in de Sitter space time, ours is a full quantum field theoretical treatment that includes coupling between all modes as befits a local quantum field theory and consistently trace over all the superhorizon degrees of freedom.

We find that the entanglement entropy asymptotically grows with the physical volume as more wavevectors cross the Hubble radius. The method is generalized to a wave packet description of single particle states and we study in detail the case of wave packets sharply localized in momentum around a wavevector $k_0 \gg H$ and localized in space on scales much smaller than the Hubble radius all throughout the near de Sitter inflationary state. We find that under these conditions, the entanglement entropy for wavepackets is approximately the same as that for plane waves and assess the corrections.

As mentioned above, the lack of kinematic thresholds implies that quanta can decay on many quanta of the same field, in particular for cubic interactions a single particle state can decay into two particles of the same field, however the decay process does not stop at the two particle level, but instead is a *cascade* decay $1 \to 2 \to 3 \to \cdots$. We provide a non-perturbative framework to study this cascade decay process and argue that for weak (cubic) coupling λ there is a hierarchy of time scales and the cascade is controlled by this weak coupling. The probability of multiparticle states is suppressed by λ^2 for each extra particle in the final state, the time scales of production and decay of multiparticle states are also separated by $1/\lambda^2$.

We comment on possible relationship with non-gaussianity, in particular pointing out the

relationship between the quantum correlations between subhorizon and superhorizon quanta from particle decay and the bispectrum of scalar perturbations in the squeezed (local) limit. Furthermore, we speculate as to whether the information "lost" as modes cross the horizon is "recovered" when the modes re-enter the horizon during the matter dominated era. This study then bridges the main concepts of entanglement between spatial regions explored in ref.[172], with momentum space entanglement and coarse graining[173] and quantum entanglement via particle decay[198] in inflationary cosmology.

5.2 QUANTUM FIELD THEORETICAL WIGNER-WEISSKOPF TREATMENT DECAY WIDTH

The method developed in refs.[175, 198, 162] is a quantum field theoretical generalization of the Wigner-Weisskopf method used in quantum optics[100, 204].

We consider a scalar field minimally coupled to gravity in a spatially flat de Sitter space-time with scale factor $a(t) = e^{Ht}$. In comoving coordinates, the action is given by

$$S = \int d^3x\, dt\, a^3(t) \left\{ \frac{1}{2}\dot{\phi}^2 - \frac{(\nabla\phi)^2}{2a^2} - \frac{M^2}{2}\phi^2 - \lambda\, \phi^3 \right\}, , \tag{5.2.1}$$

It is convenient to pass to conformal time $\eta = -e^{-Ht}/H$ with $d\eta = dt/a(t)$ and introduce a conformal rescaling of the fields

$$a(t)\phi(\vec{x}, t) = \chi(\vec{x}, \eta). \tag{5.2.2}$$

The action becomes (after discarding surface terms that will not change the equations of motion)

$$S = \int d^3x\, d\eta \left\{ \frac{1}{2}\left[\chi'^2 - (\nabla\chi)^2 - \mathcal{M}^2(\eta)\, \chi^2 \right] - \lambda\, C(\eta)\, \chi^3 \right\}, \tag{5.2.3}$$

with primes denoting derivatives with respect to conformal time η and

$$\mathcal{M}^2(\eta) = M^2 C^2(\eta) - \frac{C''(\eta)}{C(\eta)}\, , \tag{5.2.4}$$

where for de Sitter spacetime

$$C(\eta) = a(t(\eta)) = -\frac{1}{H\eta}. \tag{5.2.5}$$

In this case, the effective time dependent mass is given by

$$\mathcal{M}^2(\eta) = \Big[\frac{M^2}{H^2} - 2\Big]\frac{1}{\eta^2}\,. \tag{5.2.6}$$

The free field Heisenberg equations of motion for the spatial Fourier modes of the field with wavevector k are given by

$$\chi''_{\vec{k}}(\eta) + \Big[k^2 - \frac{1}{\eta^2}\Big(\nu^2 - \frac{1}{4}\Big)\Big]\chi_{\vec{k}}(\eta) = 0\,, \tag{5.2.7}$$

where

$$\nu^2 = \frac{9}{4} - \frac{M^2}{H^2}\,. \tag{5.2.8}$$

This can be solved to find the two linearly independent solutions of (5.2.7):

$$g_\nu(k;\eta) = \frac{1}{2}\, i^{\nu+\frac{1}{2}}\sqrt{-\pi\eta}\, H^{(1)}_\nu(-k\eta) \tag{5.2.9}$$

$$f_\nu(k;\eta) = \frac{1}{2}\, i^{-\nu-\frac{1}{2}}\sqrt{-\pi\eta}\, H^{(2)}_\nu(-k\eta) = g^*_\nu(k;\eta)\,, \tag{5.2.10}$$

where $H^{(1,2)}_\nu(z)$ are Hankel functions. Expanding the field operator in this basis yields

$$\chi(\vec{x},\eta) = \frac{1}{\sqrt{V}}\sum_{\vec{k}}\Big[a_{\vec{k}}\, g_\nu(k;\eta)\, e^{i\vec{k}\cdot\vec{x}} + a^\dagger_{\vec{k}}\; g^*_\nu(k;\eta)\, e^{-i\vec{k}\cdot\vec{x}}\Big]\,. \tag{5.2.11}$$

The Bunch-Davies vacuum is defined such that

$$a_{\vec{k}}|0\rangle = 0\,, \tag{5.2.12}$$

and the Fock space states are obtained in the usual manner, i.e. by applying creation operators $a^\dagger_{\vec{k}}$ to the vacuum.

In what follows we consider a light scalar field with $M \ll H$ and write

$$\nu = \frac{3}{2} - \Delta,\ \Delta = \frac{M^2}{3H^2} + \cdots \ll 1\,. \tag{5.2.13}$$

For light scalar fields with $\Delta \ll 1$ quantum loop corrections feature an infrared enhancement from the emission and absorption of superhorizon quanta that is manifest as poles in Δ[229,

175]. Below we exploit the expansion in Δ implemented in ref.[229, 175] to leading order, isolating the most infrared sensitive contributions to the entanglement entropy from these processes.

In the Schröedinger picture the quantum states $|\Psi(\eta)\rangle$ obey

$$i\frac{d}{d\eta}|\Psi(\eta)\rangle = H(\eta)\,|\Psi(\eta)\rangle \tag{5.2.14}$$

where in an expanding cosmology the Hamiltonian $H(\eta)$ is generally a function of η in marked contrast to the situation in Minkowski space-time, where it is constant. Introducing the time evolution operator $U(\eta,\eta_0)$ obeying

$$i\frac{d}{d\eta}U(\eta,\eta_0) = H(\eta)\,U(\eta,\eta_0), \quad U(\eta_0,\eta_0) = 1, \tag{5.2.15}$$

the solution of the Schröedinger equation is $|\Psi(\eta)\rangle = U(\eta,\eta_0)\,|\Psi(\eta_0)\rangle$. Now separate out the interaction Hamiltonian by writing $H(\eta) = H_0(\eta) + H_i(\eta)$ with $H_0(\eta)$ the non-interacting Hamiltonian, and introduce the time evolution operator of the free theory $U_0(\eta,\eta_0)$ satisfying

$$i\frac{d}{d\eta}U_0(\eta,\eta_0) = H_0(\eta)\,U_0(\eta,\eta_0), \quad i\frac{d}{d\eta}U_0^{-1}(\eta,\eta_0) = -U_0^{-1}(\eta,\eta_0)\,H_0(\eta), \quad U_0(\eta_0,\eta_0) = 1, \tag{5.2.16}$$

the interaction picture states are defined as

$$|\Psi(\eta)\rangle_I = U_I(\eta,\eta_0)|\Psi(\eta_0)\rangle_I = U_0^{-1}(\eta,\eta_0)|\Psi(\eta)\rangle. \tag{5.2.17}$$

Here $U_I(\eta,\eta_0)$ is the time evolution operator in the interaction picture and obeys

$$\frac{d}{d\eta}U_I(\eta,\eta_0) = -iH_I(\eta)U_I(\eta,\eta_0), \quad U_I(\eta_0,\eta_0) = 1 \tag{5.2.18}$$

and

$$H_I(\eta) = U_0^{-1}(\eta,\eta_0)H_i(\eta)U_0(\eta,\eta_0), \tag{5.2.19}$$

where χ is the free field Heisenberg field operator in eq.(5.2.11).

5.2.1 Transition amplitudes and probability

Now consider a cubic interaction Hamiltonian for a scalar field which we label as $\chi(\vec{x},\eta)$ after the conformal rescaling described above:

$$H_I(\eta) = -\frac{\lambda}{H\,\eta}\int d^3x\ \chi^3(\vec{x},\eta)\,. \tag{5.2.20}$$

We can then use the expansion of the scalar field χ given by (5.2.11) to compute the transition amplitude for a one particle state to decay into two particles $\chi_{\vec{k}} \to \chi_{\vec{p}} + \chi_{\vec{k}-\vec{p}}$ as depicted in fig. (24):

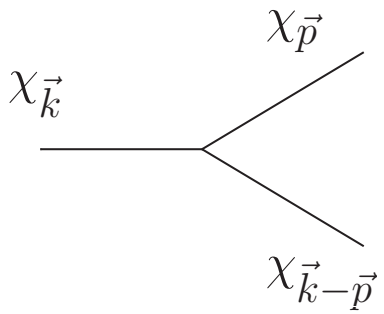


Figure 24: The decay $\chi_{\vec{k}} \to \chi_{\vec{p}} + \chi_{\vec{k}-\vec{p}}$.

$$\mathcal{A}_{\chi\to\chi\chi}(\vec{k},\vec{p};\eta) = \frac{6\,i\,\lambda}{H\,\sqrt{V}}\int_{\eta_0}^{\eta}\frac{d\eta_1}{\eta_1}\,g_\nu(k;\eta_1)\,g^*_\nu(p;\eta_1)\,g^*_\nu(|\vec{k}-\vec{p}|;\eta_1). \tag{5.2.21}$$

The total transition probability is

$$\mathcal{P}_{\chi\to\chi\chi}(k;\eta) = V\int\frac{d^3p}{(2\pi)^3}\,\left|\mathcal{A}_{\chi\to\chi\chi}(\vec{k},\vec{p};\eta)\right|^2 = \int_{\eta_0}^{\eta}d\eta_2\int_{\eta_0}^{\eta}d\eta_1\ \Sigma(k\,;\eta_1,\eta_2) \tag{5.2.22}$$

where

$$\Sigma(k\,;\eta_1,\eta_2) = \frac{36\,\lambda^2\ g^*_\nu(k,\eta_2)\,g_\nu(k,\eta_1)}{H^2\,\eta_1\,\eta_2}\int\frac{d^3p}{(2\pi)^3}\ g^*_\nu(p,\eta_1)\,g^*_\nu(q,\eta_1)\,g_\nu(p,\eta_2)\,g_\nu(q,\eta_2), \tag{5.2.23}$$

where $q = |\vec{k}-\vec{p}|$. Note that this kernel has the property that

$$\Sigma(k\,;\eta_2,\eta_1) = \Sigma^*(k\,;\eta_1,\eta_2)\,. \tag{5.2.24}$$

Introducing the identity $1 = \Theta(\eta_2 - \eta_1) + \Theta(\eta_1 - \eta_2)$ in the (conformal) time integrals and using (5.2.24) we find

$$\mathcal{P}_{\chi\to\chi\chi}(k;\eta) = 2\int_{\eta_0}^{\eta} d\eta_2 \int_{\eta_0}^{\eta_2} d\eta_1 \ \mathrm{Re}\Big[\Sigma(k\,;\eta_1,\eta_2)\Big] \tag{5.2.25}$$

from which we obtain the *transition rate* as

$$\Gamma(\eta) \equiv \frac{d}{d\eta}\mathcal{P}_{\chi\to\chi\chi}(k;\eta) = 2\int_{\eta_0}^{\eta} d\eta' \ \mathrm{Re}\big[\Sigma(k\,;\eta,\eta')\big] \tag{5.2.26}$$

In Minkowski space-time ($\eta \to t$), if the kinematics of the transition is allowed, i.e. if energy-momentum conservation obtains, the transition is to on-shell states and the transition probability grows linearly in time, exhibiting secular growth. In the long time limit the transition rate becomes a constant. This is basically how the result from Fermi's Golden rule comes about. If, on the other hand energy-momentum conservation is not fulfilled, the probability becomes constant at asymptotically long times, with a vanishing transition rate, describing virtual processes that contribute to wave function renormalization. A true decay of the quantum state is therefore reflected in secular *growth* of the transition probability and a transition rate that either remains constant or grows at asymptotically long time. In de Sitter space time the lack of a global time-like Killing vector implies the lack of kinematic thresholds. As discussed earlier in ref.[229, 175] and confirmed in ref.[194], quanta of a single field can decay into other quanta of the same field regardless of the mass of the field.

5.2.2 Wigner-Weisskopf theory in de Sitter space time:

In this subsection, we review the work in refs.[175, 198, 162], as the implementation of the quantum field theoretical Wigner-Weisskopf formulation is crucial in constructing states whose time evolution is manifestly unitary.

Expanding the interaction picture state $|\Psi(\eta)\rangle_I$ in Fock states $|n\rangle$ obtained as usual by applying the creation operators on to the (bare) vacuum state as

$$|\Psi(\eta)\rangle = \sum_n C_n(\eta)|n\rangle \tag{5.2.27}$$

the evolution of the state in the interaction picture given by eqn. (5.2.17) yields

$$i\frac{d}{d\eta}|\Psi(\eta)\rangle = H_I(\eta)|\Psi(\eta)\rangle \tag{5.2.28}$$

which in terms of the coefficients $C_n(\eta)$ become

$$\frac{d\,C_n(\eta)}{d\eta} = -i\sum_m C_m(\eta)\langle n|H_I(\eta)|m\rangle\,, \tag{5.2.29}$$

it is convenient to separate the diagonal matrix elements from those that represent transitions, writing

$$\frac{d\,C_n(\eta)}{d\eta} = -iC_n(\eta)\langle n|H_I(\eta)|n\rangle - i\sum_{m\neq n} C_m(\eta)\langle n|H_I(\eta)|m\rangle\,. \tag{5.2.30}$$

Although this equation is exact, it provides an infinite hierarchy of simultaneous equations when the Hilbert space of states $|n\rangle$ is infinite dimensional. The Wigner-Weisskopf method consists of two main ingredients: i) truncation of the hierarchy at a given order in the perturbative expansion, ii) a Markovian approximation that yields the long time asymptotics of the coefficients.

In ref.[175] the equivalence between the Wigner-Weisskopf method, the time evolution obtained from the Dyson resummation of propagators in terms of the self-energy and the dynamical renormalization group was shown in Minkowski space time. Hence this method provides a non-perturbative resummation to obtain the real time dynamics of quantum states.

We begin by implementing this program to lowest order, and provide a roadmap for implementation at arbitrary higher order in section (5.6) where we also study "cascade processes" that are available in de Sitter space time.

Thus, consider the case when a state $|A\rangle$, say, couples to a set of states $|\kappa\rangle$, which in turn couple back to $|A\rangle$ via H_I. Then to lowest order in the interaction, the system of equation closes in the form

$$\frac{d\,C_A(\eta)}{d\eta} = -i\langle A|H_I(\eta)|A\rangle\,C_A(\eta) - i\sum_{\kappa\neq A}\langle A|H_I(\eta)|\kappa\rangle\,C_\kappa(\eta) \tag{5.2.31}$$

$$\frac{d\,C_\kappa(\eta)}{d\eta} = -i\langle \kappa|H_I(\eta)|\kappa\rangle\,C_\kappa(\eta) - i\,\langle \kappa|H_I(\eta)|A\rangle\,C_A(\eta) \tag{5.2.32}$$

where the $\sum_{\kappa \neq A}$ is over all the intermediate states coupled to $|A\rangle$ via H_I representing transitions. By including the diagonal terms $\langle n|H_I(\eta)|n\rangle\, C_n$ specifically, we can also consider mass counterterms[175], however, we will neglect these terms in the sequel since we are not concerned with either mass generation or renormalization in this article.

Consider the initial value problem in which at time $\eta = \eta_0$ the state of the system is given by $|\Psi(\eta = \eta_0)\rangle = |A\rangle$ so that

$$C_A(\eta_0) = 1\ , C_{\kappa \neq A}(\eta = \eta_0) = 0. \tag{5.2.33}$$

We can then solve (5.2.32) and substitute the solution back into (5.2.31) to find

$$C_\kappa(\eta) = -i \int_{\eta_0}^{\eta} \langle \kappa|H_I(\eta')|A\rangle\, C_A(\eta')\, d\eta' \tag{5.2.34}$$

$$\frac{d\, C_A(\eta)}{d\eta} = -\int_{\eta_0}^{\eta} \Sigma(\eta, \eta')\, C_A(\eta')\, d\eta' \tag{5.2.35}$$

where

$$\Sigma(\eta, \eta') = \sum_\kappa \langle A|H_I(\eta)|\kappa\rangle \langle \kappa|H_I(\eta')|A\rangle. \tag{5.2.36}$$

This integro-differential equation with *memory* yields a non-perturbative solution for the time evolution of the amplitudes and probabilities. We can construct an approximation scheme to solve this equation as follows. First note that the time evolution of $C_A(\eta)$ as determined by eqn. (5.2.35) is *slow* in the sense that the relevant time scale is determined by a weak coupling kernel Σ. This allows us to introduce a Markovian approximation in terms of an expansion in derivatives of C_A as follows: define

$$W_0(\eta, \eta') = \int_{\eta_0}^{\eta'} \Sigma(\eta, \eta'') d\eta'' \tag{5.2.37}$$

so that

$$\Sigma(\eta, \eta') = \frac{d}{d\eta'} W_0(\eta, \eta'), \quad W_0(\eta, \eta_0) = 0. \tag{5.2.38}$$

Integrating by parts in eq.(5.2.35) we obtain

$$\int_{\eta_0}^{\eta} \Sigma(\eta, \eta')\, C_A(\eta')\, d\eta' = W_0(\eta, \eta)\, C_A(\eta) - \int_{\eta_0}^{\eta} W_0(\eta, \eta')\, \frac{d}{d\eta'} C_A(\eta')\, d\eta'. \tag{5.2.39}$$

The first term has "erased" the memory in the kernel by setting both time arguments to be the time of interest, while the second term on the right hand side is formally of *fourth order* in H_I. Integrating by parts successively as discussed in ref.[175] a systematic approximation scheme can be developed. To leading order in the coupling (second order in H_I), we will neglect the second term on the right hand side of (5.2.39), in which case eqn. (5.2.35) becomes

$$\frac{d\,C_A(\eta)}{d\eta} + W_0(\eta,\eta)\,C_A(\eta) = 0 \tag{5.2.40}$$

with solution

$$C_A(\eta) = e^{-\int_{\eta_0}^{\eta} W_0(\eta',\eta')\,d\eta'}, \quad W_0(\eta',\eta') = \int_{\eta_0}^{\eta'} \Sigma(\eta',\eta'')d\eta'' \,. \tag{5.2.41}$$

Introducing the *real quantities* $\mathcal{E}_A(\eta)$, $\Gamma_A(\eta)$ as

$$\int_{\eta_0}^{\eta'} \Sigma(\eta',\eta'')d\eta'' = i\,\mathcal{E}_A(\eta') + \frac{1}{2}\,\Gamma_A(\eta') \tag{5.2.42}$$

where

$$\Gamma_A(\eta') = 2\int_{\eta_0}^{\eta'} \mathrm{Re}\Big[\Sigma(\eta',\eta'')\Big]d\eta'' \tag{5.2.43}$$

in terms of which

$$C_A(\eta) = e^{-i\int_{\eta_0}^{\eta}\mathcal{E}_A(\eta')d\eta'}\; e^{-\frac{1}{2}\int_{\eta_0}^{\eta}\Gamma_A(\eta')d\eta'} \,. \tag{5.2.44}$$

When the state A is a single particle state, radiative corrections to the mass are extracted from $\mathcal{E}_A$ and

$$\Gamma_A(\eta) = -\frac{d}{d\eta}\ln\Big[\big|C_A(\eta)\big|^2\Big] \tag{5.2.45}$$

is identified as a (conformal) time dependent decay rate. Comparing these expressions with the transition probability (5.2.25) we see from (5.2.45) that

$$\big|C_A(\eta)\big|^2 = e^{-\mathcal{P}_{\chi\to\chi\chi}(k;\eta)} \,, \tag{5.2.46}$$

and that $\Gamma(\eta)$ is exactly the same as expression (5.2.26).

5.2.3 Unitarity

One of our main goals is to study the entanglement entropy from tracing over superhorizon degrees of freedom. Thus it is important to make sure that the loss of information encoded in the entanglement entropy is a genuine effect of the tracing procedure and not a consequence of approximations in the evolution of the quantum state. Unitarity follows from the set of equations (5.2.28), combining these with their complex conjugates it is straightforward to confirm that

$$\frac{d}{d\eta}\sum_n |C_n(\eta)|^2 = 0\,. \tag{5.2.47}$$

therefore with the initial conditions (5.2.33) it follows that

$$\sum_n |C_n(\eta)|^2 = 1\,. \tag{5.2.48}$$

Although this is an exact statement, we now show that the Wigner-Weisskopf approximation and its Markovian implementation maintain unitary time evolution.

Using (5.2.34) consider

$$\sum_\kappa |C_\kappa(\eta)|^2 = \int_{\eta_0}^{\eta} d\eta_1 C_A^*(\eta_1)\int_{\eta_0}^{\eta} d\eta_2 \Sigma(\eta_1,\eta_2) C_A(\eta_2). \tag{5.2.49}$$

Inserting $1 = \Theta(\eta_1-\eta_2)+\Theta(\eta_2-\eta_1)$ as we did earlier, it follows that

$$\begin{aligned}\sum_\kappa |C_\kappa(\eta)|^2 &= \int_{\eta_0}^{\eta} d\eta_1 C_A^*(\eta_1)\int_{\eta_0}^{\eta_1} d\eta_2 \Sigma(\eta_1,\eta_2) C_A(\eta_2) \\ &+ \int_{\eta_0}^{\eta} d\eta_2 C_A(\eta_2)\int_{\eta_0}^{\eta_2} d\eta_1 \Sigma(\eta_1,\eta_2) C_A^*(\eta_1).\end{aligned} \tag{5.2.50}$$

Using $\Sigma(\eta_1,\eta_2) = \Sigma^*(\eta_2,\eta_1)$, relabelling $\eta_1 \leftrightarrow \eta_2$ in the second line of (5.2.50) and using (5.2.35), we find

$$\begin{aligned}\sum_\kappa |C_\kappa(\eta)|^2 &= -\int_{\eta_0}^{\eta} d\eta_1 \Big[C_A^*(\eta_1)\frac{d}{d\eta_1}C_A(\eta_1) + C_A(\eta_1)\frac{d}{d\eta_1}C_A^*(\eta_1)\Big] \\ &= -\int_{\eta_0}^{\eta} d\eta_1 \frac{d}{d\eta_1}|C_A(\eta_1)|^2 = 1-|C_A(\eta)|^2\end{aligned} \tag{5.2.51}$$

where we have used the initial condition $C_A(\eta_0) = 1$. This is the statement of unitary time evolution, namely

$$|C_A(\eta)|^2 + \sum_\kappa |C_\kappa(\eta)|^2 = |C_A(\eta_0)|^2 \tag{5.2.52}$$

To leading order in the Markovian approximation, the unitarity relation becomes

$$\sum_\kappa |C_\kappa(\eta)|^2 = -2 \int_{\eta_0}^{\eta} \Big|C_A(\eta_1)\Big|^2 \, \mathrm{Re}\Big[W_0(\eta_1, \eta_1)\Big] \, d\eta_1 = 1 - |C_A(\eta)|^2 \tag{5.2.53}$$

where $C_A(\eta_0) = 1$.

5.3 PARTICLE DECAY: ENTANGLEMENT ACROSS THE HORIZON:

In the scalar theory described by eq.(5.2.20) the cubic interaction allows a single particle state $|1_{\vec{k}}\rangle$ to *decay* into two particle states $|1_{\vec{k}-\vec{p}}; 1_{\vec{p}}\rangle$[229, 175]. To lowest order in the coupling the matrix element for this process is given up to an overall phase by

$$\mathcal{M}(p; k; \eta) = \langle 1_{\vec{k}-\vec{p}}; 1_{\vec{p}} | H_I(\eta) | 1_{\vec{k}} \rangle = -\frac{6\lambda}{H\eta\sqrt{V}} \, g_\nu(k; \eta) \, g^*_\nu(p; \eta) \, g^*_\nu(|\vec{k} - \vec{p}|; \eta) \, . \tag{5.3.1}$$

Consider an initial single particle state $|1_{\vec{k}}\rangle$ at time η_0. Upon time evolution in the interaction picture this state evolves into

$$|\Psi(\eta)\rangle_I = C_k(\eta)|1_{\vec{k}}\rangle + \sum_{\vec{p}} C_p(k; \eta)|1_{\vec{k}-\vec{p}}; 1_{\vec{p}}\rangle \quad ; \quad C_k(\eta_0) = 1 \; ; \; C_p(k, \eta_0) = 0 \, . \tag{5.3.2}$$

This is an *entangled* state in which pairs of particles with momenta $\vec{k} - \vec{p}$, $\vec{p}$ *are correlated.* In particular if $\vec{k}$ is *subhorizon* and $\vec{p}$ is *superhorizon*, the quantum state (5.3.2) describes entanglement and correlation of particles across the horizon.

The coefficients in the state (5.3.2) are the solutions of the (WW) equations, namely

$$\frac{d}{d\eta} C_k(\eta) = -\int_{\eta_0}^{\eta} \Sigma(k, \eta, \eta') \, C_k(\eta') \, d\eta' \, , \tag{5.3.3}$$

$$C_p(k; \eta) = -i \int_{\eta_0}^{\eta} \mathcal{M}(p; k; \eta') \, C_k(\eta') \, d\eta' \, , \tag{5.3.4}$$

where the matrix element is given by eq.(5.3.1). We will focus on the asymptotic limit where $\eta \to 0^-; \eta_0 \to -\infty$.

The self-energy (5.2.36) is given by[1]

$$\Sigma(k,\eta,\eta') = \sum_{\vec{p}} \mathcal{M}^*(p;k;\eta)\,\mathcal{M}(p;k;\eta')\,, \tag{5.3.5}$$

were the matrix elements are given by eq.(5.3.1) leading to the result given by (5.2.23).

As discussed in detail in ref.[175], as $\Delta \to 0$ the integral features infrared divergences from regions in which the momenta are superhorizon, namely $p\eta, p\eta' \ll 1$ and $|\vec{k}-\vec{p}|\eta, |\vec{k}-\vec{p}|\eta' \ll 1$. Both of these momentum regions yield the same infrared contribution as a single pole in Δ[175], as can be seen as follows. For superhorizon modes ($-p\eta \ll 1$) the mode functions (5.2.9) behave (up to an overall phase) as

$$g_\nu(p;\eta) \simeq \frac{1}{\sqrt{2}} \frac{1}{p^{\frac{3}{2}-\Delta}\,(-\eta)^{1-\Delta}} \tag{5.3.6}$$

and for subhorizon modes $-k\eta \gg 1$

$$g_\nu(k;\eta) = \frac{1}{\sqrt{2k}}\; e^{-ik\eta}. \tag{5.3.7}$$

Therefore for $p \ll (-1/\eta)$ and $k \gg (-1/\eta)$ the matrix element (5.3.1) becomes (up to an overall phase)

$$\mathcal{M}(p,k;\eta) \simeq \frac{6\,\lambda}{2\sqrt{2}Hk\sqrt{V}\,(-\eta)^{2-\Delta}}\,\frac{1}{p^{\frac{3}{2}-\Delta}}. \tag{5.3.8}$$

The contribution to the self-energy from superhorizon modes with $p \leq \mu \lesssim (-1/\eta)$ (with μ an infrared cutoff) yields

$$\frac{V}{2\pi^2}\int_0^\mu p^2\,\mathcal{M}^*(p;k;\eta)\,\mathcal{M}(p;k;\eta')\,dp = \frac{9\lambda^2}{8\pi^2H^2k^2\eta^2\eta'^2\Delta}\left[1+\Delta\ln[\mu^2\eta\eta'] + \cdots\right]. \tag{5.3.9}$$

The processes that contributes to leading order in Δ is the emission of *superhorizon quanta*, depicted in fig. (25)

The simple rules to extract the leading order contribution in Δ are given in ref.[175], where the cancellation of the infrared regulator μ from the contributions of the subhorizon modes, for which one can safely set $\Delta = 0$, is also shown in detail. In particular, the

[1] This expression corrects a prefactor in ref.[175].

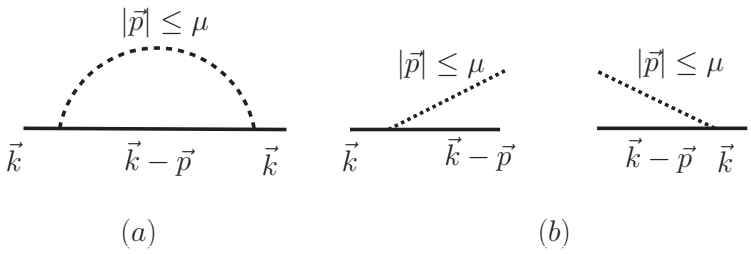


Figure 25: Processes that contribute to the leading order poles in Δ: (a) intermediate state of superhorizon modes, (b) emission and absorption of superhorizon quanta, with $\mu \lesssim (-1/\eta)$.

appendix of ref.[175] shows how the contribution of the subhorizon modes replaces the term $\ln[\mu^2\eta\eta'] \to \ln[k^2\eta\eta']$ which to leading order in Δ can be written as $1+\Delta\ln[k^2\eta\eta'] \simeq [k^2\eta\eta']^{\Delta}$. The contribution from the region $|\vec{k}-\vec{p}| \ll \mu$ yields an overall factor 2 in the self-energy so that to leading order in Δ (as can be seen by rerouting the loop momentum)

$$\Sigma(k,\eta,\eta') = \frac{\alpha}{k^{2-2\Delta}\eta^{2-\Delta}\,\eta'^{\,2-\Delta}} \quad , \quad \alpha = \frac{9\,\lambda^2}{4\pi^2\,H^2\,\Delta} = \frac{27\,\lambda^2}{4\pi^2\,M^2}\,. \tag{5.3.10}$$

Using the result in eq.(5.2.41) we finally find that to leading order,

$$C_k(\eta) = \exp\Big[-\frac{\alpha}{2\,z^{(2-2\Delta)}}\Big], \quad z = (-k\eta)\,, \tag{5.3.11}$$

where we have approximated $\alpha/2\,z_0^{2-2\Delta} \to 0$ since $-k\eta_0 \gg 1$ as the physical wavevector of the initial particle is deep inside the Hubble radius at the initial time and it is assumed to remain inside the Hubble radius during the evolution.

5.4 ENTANGLEMENT ENTROPY

The pure state density matrix corresponding to the entangled state of eq.(5.3.2) is

$$\rho(\eta) = |\Psi(\eta)\rangle\langle\Psi(\eta)|. \tag{5.4.1}$$

Now let us trace over the superhorizon physical wavevectors $-\vec{p}\eta \lesssim 1$. This leads us to the *mixed state* density matrix for modes whose wavelengths are *inside* the horizon during the evolution

$$\rho_r(\eta) = |C_k(\eta)|^2 |1_{\vec{k}}\rangle\langle 1_{\vec{k}}| + 2\sum_{-p\eta\lesssim 1} |C_p(k;\eta)|^2 |1_{\vec{k}-\vec{p}}\rangle\langle 1_{\vec{k}-\vec{p}}| \tag{5.4.2}$$

where the factor 2 accounts for the two regions of superhorizon (physical) momenta $-p\eta < 1$ and $-|\vec{k}-\vec{p}|\eta < 1$ which yield the same contribution, as can be easily seen after a relabelling of momenta.

The entanglement entropy is the Von-Neumann entropy for the reduced density matrix, we find

$$S(\eta) = -n_k(\eta)\ln n_k(\eta) - 2\sum_{p\lesssim(-1/\eta)} n_p(\eta)\ln n_p(\eta) \tag{5.4.3}$$

where the occupation numbers of the initial and *produced* quanta are given by

$$n_k(\eta) = \langle\Psi(\eta)|a^\dagger_{\vec{k}}\, a_{\vec{k}}|\Psi(\eta)\rangle = |C_k(\eta)|^2,\;\; n_p(\eta) = \langle\Psi(\eta)|a^\dagger_{\vec{p}}\, a_{\vec{p}}|\Psi(\eta)\rangle = |C_p(k;\eta)|^2\,. \tag{5.4.4}$$

Note that the unitarity relation in eq.(5.2.53) implies that

$$\sum_{\vec{p}} n_p(\eta) = 1 - n_k(\eta)\,. \tag{5.4.5}$$

as expected on physical grounds.

At this stage it is important to highlight how unitarity is manifest to leading order in the Δ expansion as this feature simplifies the calculation of the entanglement entropy considerably. The main point is that the unitarity relation (5.2.53) implies that the contribution of superhorizon modes is the dominant one. This can be seen clearly from the following arguments: to leading order in Δ we can neglect the term 2Δ in the exponent of z in the

solution (5.3.11) and in the term $-\Delta$ in the exponent of $(-\eta)$ in the matrix element (5.3.8) for superhorizon modes. Now the coefficient

$$C_p(k;\eta) = -i\int_{\eta_0}^{\eta} \mathcal{M}(p;k;\eta_1)\, C_k(\eta_1)\, d\eta_1 \simeq -i\frac{2\pi}{\sqrt{2V}}\frac{\sqrt{\Delta}}{p^{\frac{3}{2}-\Delta}}\int_{y(\eta_0)}^{y(\eta)} e^{-y^2/2}dy \tag{5.4.6}$$

where we used the definition of α given by eqn. (5.3.10) and changed variables of integration to $\eta_1 = \sqrt{\alpha}/k\, y$. This expression clearly exhibits that the contribution to $|C_p(k;\eta)|^2$ from superhorizon modes, to leading order in Δ can be written in the following factorized form:

$$|C_p(k;\eta)|^2 = F[k;\eta]\ \frac{\Delta}{V\, p^{3-2\Delta}}\,. \tag{5.4.7}$$

The dependence on Δ is a manifestation of unitarity to leading order; if we compute the integral in eq.(5.4.7) over superhorizon modes

$$\sum_{p\lesssim(-1/\eta)} |C_p(k;\eta)|^2 = \frac{F[k;\eta]\,\Delta}{2\pi^2}\int_0^{(-1/\eta)}\frac{p^2dp}{p^{3-2\Delta}} = \frac{F[k;\eta]}{4\pi^2}\,(-1/\eta)^{2\Delta}, \tag{5.4.8}$$

the Δ in the numerator in eq.(5.4.7) cancels the single pole in Δ from the integral giving an $\mathcal{O}(1)$ contribution, which is what is necessary to satisfy the unitarity condition (5.2.53) to leading order in Δ.

This result is similar to that found in the case of particle decay in Minkowski space time[198]: in this case the particles produced from the decay of a parent particle feature a Lorentzian distribution in energy, with width Γ the decay width of the parent particle and amplitude $1/\Gamma$, so that the energy integral over the distribution is $\mathcal{O}(1)$. In ref.[198] it is proven to leading order in the perturbative expansion $\mathcal{O}(\Gamma)$ that this narrow distribution of large amplitude is the main reason for the fulfillment of unitarity to leading order in the Wigner-Weisskopf approximation. In the case of de Sitter space time, the distribution function of the particles produced with superhorizon wavevectors is $\propto \Delta/p^{3-2\Delta}$ whose momentum integral over the region of superhorizon momenta is also of $\mathcal{O}(1)$.

Thus in the limit $\Delta \ll 1$ the sum $\sum_p |C_p(\eta)|^2$ is dominated by the superhorizon momenta and from the unitarity relation (5.2.53) we find

$$\mathrm{Tr}\rho_r(\eta) = |C_k(\eta)|^2 + \sum_p |C_p(\eta)|^2 = 1\,. \tag{5.4.9}$$

Although the integral in $F[k;\eta]$ can be written in terms of error functions, the unitarity relation (5.2.53) and the result (5.4.9) furnish a more direct evaluation. Consider

$$\begin{aligned}\sum_p |C_p(k;\eta)|^2 &= \int_{\eta_0}^{\eta}\int_{\eta_0}^{\eta}\sum_p \mathcal{M}^*(p;k;\eta_1)\,\mathcal{M}(p;k;\eta_2)\,C_k^*(\eta_1)\,C_k(\eta_2)\,d\eta_1\,d\eta_2 \\ &= \int_{\eta_0}^{\eta}\int_{\eta_0}^{\eta}\Sigma(k,\eta_1,\eta_2)\,C_k^*(\eta_1)C_k(\eta_2)\,d\eta_1\,d\eta_2\,. \end{aligned} \tag{5.4.10}$$

This is the same expression as in eqn. (5.2.49), so that implementing the same steps as in eqns. (5.2.50,5.2.51) leads to the unitarity relation (5.2.53), namely

$$\sum_p |C_p(k;\eta)|^2 = 1 - |C_k(\eta)|^2\,. \tag{5.4.11}$$

To leading order in Δ, the sum is dominated by the superhorizon contributions from both regions of integrations $p \lesssim (-1/\eta)$, $|\vec{k}-\vec{p}| \lesssim (-1/\eta)$ contributing equally, hence

$$\sum_{p\lesssim(-1/\eta)} |C_p(k;\eta)|^2 \simeq \frac{1}{2}\big[1-|C_k(\eta)|^2\big]\,. \tag{5.4.12}$$

Then the factorized form (5.4.7) for superhorizon modes, combined with eqn. (5.4.12) leads to

$$F[k;\eta] = \frac{2\pi^2}{(-\eta)^{-2\Delta}}\,\big[1-|C_k(\eta)|^2\big]\,, \tag{5.4.13}$$

and for $-k\eta \gg 1$ and $-p\eta \ll 1$ we find to leading order in Δ

$$|C_p(k;\eta)|^2 = \frac{2\pi^2\,\Delta}{V\,p^3\,(-p\eta)^{-2\Delta}}\,\big[1-|C_k(\eta)|^2\big]\,; \tag{5.4.14}$$

the same result is valid in the region $-k\eta \gg 1$ with $-|\vec{k}-\vec{p}|\eta \ll 1$ by replacing $p \leftrightarrow |\vec{k}-\vec{p}|$.

The long wavelength limit of eq.(5.4.14) requires a careful treatment. Since $|C_p(\eta)|^2 = n_p(\eta)$ is the distribution function of particles, for a fixed volume V there is an infrared divergence in the occupation as $p \to 0$. However, our goal is to trace over the superhorizon quanta from the decay since the initial conformal time $-\eta_0$ up to conformal time $\eta \to 0^-$. This entails that the lower momentum cutoff is determined by the mode that just becomes superhorizon at the initial time, namely

$$p_m = -1/\eta_0\,. \tag{5.4.15}$$

Now the calculation of the entanglement entropy is straightforward: let us consider

$$I = \sum_{(-1/\eta_0)\leq p\leq(-1/\eta)} |C_p(k;\eta)|^2 \ln\left[|C_p(k;\eta)|^2\right] \equiv I_1 + I_2 \tag{5.4.16}$$

with

$$\begin{aligned} I_1 &= \left[1-|C_k(\eta)|^2\right] \ln\left[\frac{2\pi^2\Delta(-\eta_0)^3}{V}\left[1-|C_k(\eta)|^2\right]\right] \Delta \int_{(-1/\eta_0)}^{(-1/\eta)} (-p\eta)^{2\Delta}\frac{dp}{p} \\ &= \frac{1}{2}\left[1-|C_k(\eta)|^2\right] \ln\left[\frac{2\pi^2\Delta(-\eta_0)^3}{V}\left[1-|C_k(\eta)|^2\right]\right] \left[1-x_m^{2\Delta}\right] \end{aligned} \tag{5.4.17}$$

where we have introduced

$$x_m = \frac{\eta}{\eta_0} \tag{5.4.18}$$

and changing integration variable to $x = -p\eta$

$$\begin{aligned} I_2 &= -\left[1-|C_k(\eta)|^2\right] \Delta \int_{x_m}^{1} x^{2\Delta-1} \ln\left[\frac{x^{3-2\Delta}}{x_m^3}\right] dx \\ &= \frac{1}{2}\left[1-|C_k(\eta)|^2\right]\left\{3\ln[x_m] + \frac{3-2\Delta}{2\Delta}\left[1-(x_m)^{2\Delta}\right] - 2\Delta\,(x_m)^{2\Delta}\,\ln[x_m]\right\} \end{aligned} \tag{5.4.19}$$

It is now clear that we can set $x_m \to 0$ safely in I_1 and in the terms that *do not feature poles in* Δ in I_2. The terms in I_2 that feature the $\ln[x_m]$ and the (single) pole in Δ, namely $(3/2\Delta)\times[1-(x_m)^{2\Delta}]$ yield the leading contribution for $\Delta, x_m \ll 1$.

Therefore for $\Delta \ll 1$ and $x_m \ll 1$ we find for the entanglement entropy to leading order

$$\begin{aligned} S(\eta) &\simeq \frac{\alpha}{(k\eta)^2}\, e^{-\frac{\alpha}{(k\eta)^2}} - \left[1-e^{-\frac{\alpha}{(k\eta)^2}}\right] \ln\left[1-e^{-\frac{\alpha}{(k\eta)^2}}\right] \\ &+\tfrac{1}{2}\left[1-e^{-\frac{\alpha}{(k\eta)^2}}\right]\left\{3\ln\left[\frac{a_iH_i}{a_0H_0}\right] + \ln\left[\frac{1}{2\pi^2\,\Delta}\right] + \frac{3}{2\Delta}\left[Z[\eta]-1+e^{-Z[\eta]}\right]\right\} \end{aligned} \tag{5.4.20}$$

where

$$Z[\eta] = 2\Delta\ \ln\left[\frac{\eta_0}{\eta}\right], \tag{5.4.21}$$

α is given in eqn. (5.3.10) and we have set $-\eta_0 = 1/(a_iH_i)$ and $V = 1/(a_0H_0)$ with a_i, a_0 the scale factor, and H_i, H_0 the values of the Hubble parameter at the beginning of inflation (i) and today (0) respectively, taking the physical volume today to be the Hubble volume, therefore $a_iH_i/a_0H_0 \simeq 1$. The function $Z-1+e^{-Z}$ is manifestly (semi) positive and

monotonically increasing, behaving as $\simeq Z^2/2$ for $Z \ll 1$ and as $\simeq Z$ for $Z \gg 1$. As $\eta \to 0$ the entanglement entropy grows monotonically during the time evolution.

We can $Z[\eta]$ in terms of the number of e-folds since the beginning of inflation as

$$Z[\eta] \simeq 40 \frac{M^2}{H^2} \left[1 + (N_e(\eta) - N_T)/N_T\right], \tag{5.4.22}$$

where $N_e(\eta)$ is the number of e-folds during inflation at (conformal) time η and $N_T \simeq 60$ is the total number of e-folds of the inflationary stage.

5.5 WAVE PACKETS:

The discussion above treated the initial and product particles in terms of plane waves. However, given the existence of a horizon and the intricacies that can give rise to for non-localized states, we now generalize the treatment to the case of wave packets. Quantization in a finite volume V is used throughout. Fock states describing single particle plane wave states of momentum $\vec{k}$, $|1_{\vec{k}}\rangle$, are normalized such that

$$\langle 1_{\vec{k}} | 1_{\vec{k}'} \rangle = \delta_{\vec{k},\vec{k}'} \,. \tag{5.5.1}$$

Localized single particle states are constructed as linear superpositions

$$\overline{|\vec{k}_0, \vec{x}_0\rangle} = \sum_{\vec{k}} C(\vec{k}; \vec{k}_0; \vec{x}_0) \, |1_{\vec{k}}\rangle \tag{5.5.2}$$

where $C_(\vec{k}; \vec{k}_0; \vec{x}_0)$ is the amplitude, normalized so that

$$\overline{\langle \vec{k}_0, \vec{x}_0|} \, \overline{|\vec{k}_0, \vec{x}_0\rangle} = \sum_{\vec{k}} |C(\vec{k}; \vec{k}_0; \vec{x}_0)|^2 = 1 \,. \tag{5.5.3}$$

For a monochromatic plane wave $C(\vec{k}; \vec{k}_0; \vec{x}_0) = \delta_{\vec{k},\vec{k}_0}$. The spatial wave function corresponding to the wave packet is given by

$$\Upsilon(\vec{x}) = \frac{1}{\sqrt{V}} \sum_{\vec{k}} C(\vec{k}; \vec{k}_0; \vec{x}_0) \, e^{-i\vec{k}\cdot\vec{x}} \,. \tag{5.5.4}$$

The normalization (5.5.3) implies

$$\int d^3x |\Upsilon(\vec{x})|^2 = 1\,. \tag{5.5.5}$$

For a monochromatic plane wave it follows that $\Upsilon(\vec{x})$ is a volume normalized plane wave.

The total number of particles and average momentum of the wave packet are given by

$$N(\vec{k}_0, \vec{x}_0) = \overline{\langle \vec{k}_0, \vec{x}_0|} \sum_{\vec{k}} a^\dagger_{\vec{k}} a_{\vec{k}} \overline{|\vec{k}_0, \vec{x}_0\rangle} = \sum_{\vec{k}} |C(\vec{k}; \vec{k}_0; \vec{x}_0)|^2 = 1 \tag{5.5.6}$$

and

$$\overline{\langle \vec{k}_0, \vec{x}_0|} \sum_{\vec{k}} \vec{k}\, a^\dagger_{\vec{k}} a_{\vec{k}} \overline{|\vec{k}_0, \vec{x}_0\rangle} = \sum_{\vec{k}} \vec{k} |C(\vec{k}; \vec{k}_0; \vec{x}_0)|^2 \tag{5.5.7}$$

respectively, where $a^\dagger_{\vec{k}}; a_{\vec{k}}$ are the creation and annihilation operators. If $\vec{k}_0$ is identified with the average momentum of the wave packet we assume that

$$C(\vec{k}; \vec{k}_0; \vec{x}_0) = C(\vec{k} - \vec{k}_0; \vec{x}_0)\,, \tag{5.5.8}$$

and the isotropy of $|C(\vec{k}; \vec{0}, \vec{x}_0)|^2$.

As a specific example we consider Gaussian wave packets,

$$C(\vec{k} - \vec{k}_0; \vec{x}_0) = \left[\frac{8\,\pi^{\frac{3}{2}}}{\sigma^3\, V}\right]^{\frac{1}{2}} e^{-\frac{(\vec{k}-\vec{k}_0)^2}{2\sigma^2}}\; e^{i(\vec{k}-\vec{k}_0)\cdot\vec{x}_0}\,, \tag{5.5.9}$$

where σ is the localization in momentum space. The spatial wave function is

$$\Upsilon(\vec{x}) = \left[\frac{\sigma}{\sqrt{\pi}}\right]^{3/2} e^{-i\vec{k}_0\cdot\vec{x}}\, e^{-\frac{\sigma^2}{2}(\vec{x}-\vec{x}_0)^2}\,. \tag{5.5.10}$$

The spatial wave function is localized at $\vec{x}_0$ with localization length $1/\sigma$ and the momentum wave function is localized at $\vec{k}_0$ which is the average momentum in the wave packet and the momentum localization scale is σ. The plane wave limit is obtained by formally identifying $\sigma/\sqrt{\pi} \to 1/V^{1/3}$; $V \to \infty$.

In terms of these wave functions the overlap of two wave packets with different momenta localized at different spatial points is

$$\overline{\langle \vec{q}_0; \vec{x}_0|} \overline{|\vec{k}_0; \vec{x}_0\rangle} = e^{-\frac{(\vec{k}_0-\vec{q}_0)^2}{4\sigma^2}}\,. \tag{5.5.11}$$

In the limit $\sigma \to 0$ the overlap becomes a Kronecker delta, and in particular for $k_0, q_0 \gg \sigma$ it follows that the wavepackets are nearly orthogonal since the overlap is non-vanishing for $\Delta k = k_0 - q_0 \sim \sigma$ so that $\Delta k/k_0 \ll 1$.

From the identity (5.5.8) we can infer the following important property of these wave packets which will be useful below:

$$\sum_{\vec{k}} C(\vec{k} - \vec{k}_0; \vec{x}_0)\, |1_{\vec{k}-\vec{q}}\rangle = |\overline{\vec{k}_0 - \vec{q}; \vec{x}_0\rangle}\,. \tag{5.5.12}$$

Although this result is evident with the Gaussian wave packets (5.5.9) it is quite general for localized functions of $\vec{k} - \vec{k}_0$.

The wave packet description is easily incorporated into the Wigner-Weisskopf approach to the description of the full time evolution of the quantum state of the decaying parent particle. The interaction picture quantum state (5.2.27) is generally written as

$$|\Psi(\eta)\rangle = \sum_{\vec{k}} C(\vec{k}, \vec{k}_0; \vec{x}_0; \eta)|1_{\vec{k}}\rangle + \sum_{\kappa} \mathcal{C}_\kappa(\eta)|\kappa\rangle \tag{5.5.13}$$

where the states $|\kappa\rangle$ are multiparticle states, with the initial conditions

$$C(\vec{k}; \vec{k}_0; \vec{x}_0; \eta_0) = C(\vec{k} - \vec{k}_0; \vec{x}_0) \quad ; \quad \mathcal{C}_\kappa(t = 0) = 0\,, \tag{5.5.14}$$

where $C(\vec{k} - \vec{k}_0; \vec{x}_0)$ describe the localized wave packet of the single particle state at the initial time, for example (5.5.9).

Generalizing the state (5.3.2) describing the time evolved state to lowest order in λ, to a wave packet localized at the origin in space with the gaussian profile (5.5.9), we can write

$$|\Psi(\eta)\rangle_I = \sum_{\vec{k}} C_1(\vec{k} - \vec{k}_0; \vec{0}; \eta)|1_{\vec{k}}\rangle + \sum_{\vec{p},\vec{k}} C_2(\vec{k}, \vec{p}, \vec{k}_0; \eta)|1_{\vec{k}-\vec{p}}; 1_{\vec{p}}\rangle\,, \tag{5.5.15}$$

with the initial condition

$$C_1(\vec{k} - \vec{k}_0; \vec{0}; \eta_0) = C(\vec{k} - \vec{k}_0; \vec{0}) \quad ; \quad C_2(\vec{k}, \vec{p}, \vec{k}_0; \eta_0) = 0 \tag{5.5.16}$$

with $C(\vec{k} - \vec{k}_0; \vec{0})$ given by (5.5.9).

Recall that our goal in this article was to obtain the entanglement entropy associated with the decay of single particle states with sub-Hubble physical momenta all throughout

the inflationary stage, assuming that near de Sitter inflation lasts a finite time. Namely the physical wavelength of the single particle state is always deep within the Hubble radius during the evolution. A wave packet description of single particle states, therefore must be in terms of wave packets whose physical spatial localization scale is always much smaller than the Hubble radius. Hence, we will consider wavepackets that are i) sharply localized in comoving momentum with an average momentum $\vec{k}_0$ with $k_0 \gg H; k_0 \gg \sigma$, the latter condition ensuring a sharp localization around k_0, and ii) with comoving spatial localization scale $1/\sigma \lesssim 1/H$ so that the wavepacket is localized well within the Hubble radius. Namely the condition for the wavepacket to describe single particle states with a sharp localization in momentum and with spatial localization length scale smaller than or of the order of the Hubble radius implies the following constraint:

$$k_0 \gg \sigma \gtrsim H \,. \tag{5.5.17}$$

Furthermore, consistency in tracing over degrees of freedom with super-Hubble *physical* wavelengths requires that the wavepacket is mainly composed of components with comoving momenta corresponding to physical wavelengths that are always inside the Hubble radius throughout the near de Sitter stage. This condition requires $-k_0\eta \gg -\sigma\eta \gg 1$ so that components of the wavepacket with super Hubble physical wavelengths are exponentially suppressed.

The Wigner Weisskopf method follows the steps described in detail above. The interaction Hamiltonian connects the *single particle plane wave states* $|1_{\vec{k}}\rangle$ with the two-particle plane wave states $|1_{\vec{k}-\vec{p}}; 1_{\vec{p}}\rangle$ with matrix elements given by (5.3.1) leading to the set of equations

$$\frac{d}{d\eta} C_1(\vec{k}-\vec{k}_0;\vec{0};\eta) = -\int_{\eta_0}^{\eta} \Sigma(k,\eta,\eta')\, C_1(\vec{k}-\vec{k}_0;\vec{0};\eta')\, d\eta' \,, \tag{5.5.18}$$

$$C_2(\vec{k},\vec{p},\vec{k}_0;\eta) = -i\int_{\eta_0}^{\eta} \mathcal{M}(p;k;\eta')\, C_1(\vec{k}-\vec{k}_0;\vec{0};\eta')\, d\eta' \,. \tag{5.5.19}$$

Implementing the Markovian approximation as in the plane wave case with the initial conditions (5.5.16) we find

$$C_1(\vec{k}-\vec{k}_0;\vec{0};\eta) = C_1(\vec{k}-\vec{k}_0;\vec{0};\eta_0)\, C_k(\eta) \;\; ; \;\; C_2(\vec{k},\vec{p},\vec{k}_0;\eta) = C_1(\vec{k}-\vec{k}_0;\vec{0};\eta_0)\, C_p(k;\eta) \,, \tag{5.5.20}$$

where $C_k(\eta); C_p(k;\eta)$ are the solutions of the Wigner-Weisskopf equations *for plane waves*, given by (5.3.11,5.3.4).

To obtain the reduced density matrix we would need to carry out the integration over the wavepacket variable $\vec{k}$. The wave packet profile (as function of comoving wavevectors) is chosen to be sharply peaked at $\vec{k}_0$ with a width $\sigma \ll k_0$. Therefore upon integration we can Taylor expand the integrand around $\vec{k} = \vec{k}_0$ and integrate term by term in the Taylor expansion in $\vec{k}-\vec{k}_0$, because the wavepacket profile is a function of $|\vec{k}-\vec{k}_0|$ it follows that the corrections are a series in $\sigma^2/k_0^2 \ll 1$. An example of a quantity that must be integrated in $\vec{k}$ are the matrix elements (5.3.1), which upon being integrated with the wavepacket profile can be simply written as $\mathcal{M}(p;k_0;\eta) + \mathcal{O}(\sigma^2/k_0^2) + \cdots$. The same argument applies to the coefficients

$$\begin{aligned} C_1(\vec{k}-\vec{k}_0;\vec{0};\eta) &= C(\vec{k}-\vec{k}_0;\vec{0})\, C_{k_0}(\eta) + \mathcal{O}(\sigma^2/k_0^2) + \cdots \\ C_2(\vec{k},\vec{p},\vec{k}_0;\eta) &= C(\vec{k}-\vec{k}_0;\vec{0})\, C_p(k_0;\eta) + \mathcal{O}(\sigma^2/k_0^2) + \cdots \end{aligned} \tag{5.5.21}$$

Therefore, to leading order in σ^2/k_0^2 the reduced density matrix becomes

$$\begin{aligned} \rho_r(\eta) = |C_{k_0}(\eta)|^2 \sum_{\vec{k}} \left(C(\vec{k}-\vec{k}_0;\vec{0})|1_{\vec{k}}\rangle \right) \sum_{\vec{k}'} \left(C^*(\vec{k}'-\vec{k}_0;\vec{0})\langle 1_{\vec{k}'}| \right) + \\ 2 \sum_{(-1/\eta_0)<p<(-1/\eta)} |C_p(k_0;\eta)|^2 \sum_{\vec{k},\vec{k}'} \left(C(\vec{k}-\vec{k}_0;\vec{0})|1_{\vec{k}-\vec{p}}\rangle \right) \left(C^*(\vec{k}'-\vec{k}_0;\vec{0})\langle 1_{\vec{k}'-\vec{p}}| \right) \end{aligned} \tag{5.5.22}$$

We emphasize that the trace over the superhorizon modes leading to the reduced density matrix (5.5.22) has been carried out in the orthonormal *plane wave basis*.

Using the definition of the wavepacket single particle states (5.5.2) and the property (5.5.12) we finally find to leading order in $\sigma^2/k_0^2 \ll 1$

$$\rho_r(\eta) = |C_{k_0}(\eta)|^2 \, \overline{|\vec{k}_0,\vec{0}\rangle} \; \overline{\langle \vec{k}_0,\vec{0}|} + 2 \sum_{(-1/\eta_0)<p<(-1/\eta)} |C_p(k_0;\eta)|^2 \, \overline{|\vec{k}_0-\vec{p},\vec{0}\rangle} \; \overline{\langle \vec{k}_0-\vec{p},\vec{0}|}. \tag{5.5.23}$$

For $k_0 \gg p, \sigma$ the wave-packet states $\overline{|\vec{k}_0-\vec{p},\vec{0}\rangle}$ contain plane wave components with subhorizon momenta $\simeq \vec{k}_0 - \vec{p}$ since components with wavevectors that are very different from this value are exponentially suppressed. Therefore these wavepacket states are very nearly plane wave states with subhorizon momenta $k_0 \gg -1/\eta$.

Therefore, to leading order in σ^2/k_0^2, the reduced density matrix in terms of the wave packet single particle states features the same form as for the plane wave case with the only modification being the replacement of the single particle Fock states by the localized wavepacket states of single particles. As a corollary, to leading order in σ^2/k_0^2 the entanglement entropy is the *same* either for localized wavepackets or plane waves.

The logarithmic dependence of the entanglement entropy (5.4.20) on the volume factor has a clear statistical interpretation independent of whether the description is in terms of localized wavepackets or plane wave states. Consider a dilute gas of particles whose statistical distribution or phase space density is f_p. The total *density* of particles is

$$\frac{N}{V} = \int \frac{d^3p}{(2\pi)^3}\, f_p \tag{5.5.24}$$

and the Von-Neumann entropy of this (dilute) gas is

$$S_{VN} = -\sum_p f_p \ln[f_p] = -V \int \frac{d^3p}{(2\pi)^3}\, f_p\, \ln[f_p]\,. \tag{5.5.25}$$

If the number of particles remains finite in the large volume limit, namely if the particle density scales $\propto 1/V$ in this limit, then it follows that $f_p \propto 1/V$. On the contrary, if f_p is independent of the volume as in the cases of the Maxwell-Boltzmann, Bose-Einstein or Fermi-Dirac distributions, the total density is *finite* in the infinite volume limit and the entropy is extensive. For a finite number of particles (vanishing particle density in the infinite volume limit) $f_p \propto 1/V$ and the Von-Neumann entropy is *not* extensive,

$$S_{VN} \propto N \ln[V]\,. \tag{5.5.26}$$

This is *precisely* the origin of the logarithmic dependence on the volume of the entanglement entropy: the initial state has one particle within a Hubble volume and the final state has one (of each) daughter particle, the distribution function of the daughter particles at asymptotically long times after the decay of the parent particle is $|\mathcal{C}_{\chi\psi}(p,\infty)|^2 \propto 1/V$ the inverse volume dependence is the statement that there is a finite number of particles distributed in phase space[2]. Obviously this volume dependence is independent of whether the states

[2]In the first reference in [173], only the coupling was kept in the $\ln|C(k)|$ and terms that feature a volume dependence in $|C(k)|$ were discarded as subleading. This explains a discrepancy in the logarithmic volume dependence between this ref. and our results.

are described by plane waves or wave packets, but is a statement of the simple fact that the number of particles in the volume $V = (-1/\eta_0)^3$ is finite. The dependence on the scale factor reflects the fact that more modes are crossing the Hubble radius, but the total number of particles described by these modes is still finite.

5.6 CASCADE PROCESSES: THE WAY FORWARD

In the previous section we implemented the Wigner Weisskopf method to lowest order in λ^2, but the method itself is much more general. It relies on a perturbative expansion, a truncation of the hierarchy at a given order in this expansion, and a resummation of the resulting self-energy terms that yield the long time asymptotics. For example, in quantum optics it has been implemented to study the cascade decay of many level atoms[204, 212]. As shown in[175] this resummation is a real time version of the Dyson resummation of self-energies and is equivalent to a dynamical renormalization group resummation of secular terms.

In this section we set up a roadmap to study higher order processes and along the way we exhibit the relation between the Wigner-Weisskopf method and the resummation of self-energy diagrams and a diagrammatic expansion. Given the discussion on wavepackets in the previous section, we will restrict ourselves to treating the plane wave case.

The lack of kinematic thresholds in inflationary cosmology implies that the decay of quanta occur in a *cascade* process. For example with a cubic interaction as studied above, a state with a single quanta can decay into a state with two other quanta, in turn each one of the quanta in this state can decay into two other quanta, therefore a single particle state will decay via a "cascade": $1 \to 2 \to 3 \to 4 \cdots$ depicted in fig.(26).

Each branch of the cascade corresponds to an interaction vertex and another power of the coupling, showing that the branches of the cascade are suppressed in perturbation theory. For example, the amplitude for 3 particles is down by a factor of λ (trilinear coupling) with respect to the two particle one, the four particle state is suppressed by another power of λ,

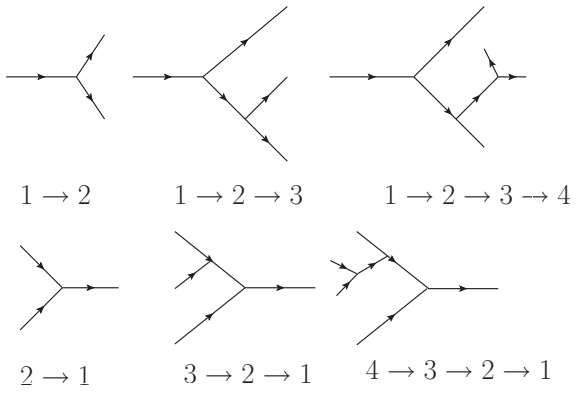


Figure 26: Upper diagrams: cascade decay $1 \to 2 \to 3 \to 4$, each vertex corresponds to a matrix element $M_{ij} \propto \lambda$. Lower diagrams: inverse processes, each vertex corresponds to the matrix element $M_{ji} = M^*_{ij} \propto \lambda$. Vacuum disconnected diagrams are neglected.

etc.

To simplify notation, let us define matrix elements that connect a state of i quanta with a state of j quanta via the interaction Hamiltonian H_I,

$$M_{ij}(\eta) = \langle [i]|H_I(\eta)|[j]\rangle \propto \lambda\,. \tag{5.6.1}$$

Here $[i], [j]$ describes the set of i, j quanta with different values of momenta. As studied above, we see that a state with a single quanta of (comoving) momentum $\vec{k}$ is connected via H_I to a state with two quanta, with momenta $\vec{q}, \vec{k} - \vec{q}$ respectively. The matrix element for this process is $\langle [1]|H_I(\eta)|[2]\rangle = \langle 1_{\vec{k}}|H_I(\eta)|1_{\vec{q}}; 1_{\vec{k}-\vec{q}}\rangle$ where the set of values $\vec{q}$ defines the two-quanta states [2]. Thus the generic matrix elements between single quanta states and two quanta states in this set are $\langle [1]|H_I(\eta)|[2]\rangle \equiv M_{12}(\eta) \propto \lambda$. The *inverse* process $[j] \to [i]$ is described by the matrix element $\langle [j]|H_I(\eta)|[i]\rangle = M_{ji}(\eta) = M^*_{ij}(\eta)$ because the Hamiltonian is hermitian.

In what follows we will only consider *connected* diagrams or processes, neglecting disconnected diagrams which do not describe transitions but rather a renormalization of the vacuum state (for discussions see[175]). Consider the cascade decay of a single particle state

$|1_{\vec{k}}\rangle$ into three particles, along with their inverse processes, neglecting the disconnected (vacuum) diagrams the typical sequence is shown in fig. (26) and the quantum state is given by

$$|\Psi(\vec{k},\eta)\rangle = C_1(k,\eta)|1_{\vec{k}}\rangle + \sum_{\vec{p}} C_2(\vec{k},\vec{p};\eta)|1_{\vec{k}-\vec{p}};1_{\vec{p}}\rangle + \sum_{\vec{p},\vec{q}} C_3(\vec{k},\vec{p},\vec{q};\eta)|1_{\vec{k}-\vec{p}};1_{\vec{q}};1_{\vec{p}-\vec{q}}\rangle + \cdots \tag{5.6.2}$$

The set of Wigner-Weisskopf equations are obtained straightforwardly as in the previous section. An important aspect in obtaining these equations is that a particular state with n particles with a fixed set of momenta has branched out from one "ancestor state", whereas it branches forward into an $n+1$ particle state where the new particle has an arbitrary momentum that is summed over. As an example of this pattern consider the 3 particle state $|1_{\vec{k}-\vec{p}};1_{\vec{q}};1_{\vec{p}-\vec{q}}\rangle$ for a fixed value of $\vec{p}$ and $\vec{q}$ (the value of $\vec{k}$ is fixed by the initial state). This state branched out from the two particle state $|1_{\vec{k}-\vec{p}};1_{\vec{p}}\rangle$ (up to relabelling the momenta and indistinguishability of the particle states), therefore *it only has one "ancestor"* as a consequence of momentum conservation. However, it branches out to 4 particle states of the form $|1_{\vec{k}-\vec{p}};1_{\vec{q}};1_{\vec{l}};1_{\vec{p}-\vec{q}-\vec{l}}\rangle$ where the wavevector $\vec{l}$ must be summed over.

The hierarchy of Wigner-Weisskopf equations reads in shortened notation

$$\begin{aligned}
\dot{C}_1(\eta) &= -i\sum_{[2]} M_{12}(\eta)C_{[2]}(\eta) && (5.6.3)\\
\dot{C}_2(\eta) &= -iM_{21}(\eta)C_1(\eta) - i\sum_{[3]} M_{23}(\eta)C_{[3]}(\eta) && (5.6.4)\\
\dot{C}_3(\eta) &= -iM_{32}(\eta)C_2(\eta) - i\sum_{[4]} M_{34}(\eta)C_{[4]}(\eta) && (5.6.5)\\
\vdots &= \vdots
\end{aligned}$$

The labels without brackets in the coefficients C_n correspond to a particular state of $n-$*particles* with a fixed set of momenta compatible with total momentum conservation whereas the sums over $[n]$ are over the n-particle states compatible with the set of wavenumbers determined by momentum conservation. The terms shown in the hierarchy (5.6.3,5.6.4,5.6.5) are the ones depicted in fig. (26) and their inverse processes: if the Hamiltonian connect the states $[i]$ with the states $[j]$ it also connects $[j]$ back with $[i]$, these are the inverse processes depicted in fig. (26).

The two terms in eqns. (5.6.4,5.6.5) have an illuminating interpretation. The first terms correspond to the "population gain" of the states with two and three particles from the decay of their ancestors states with one and two particles respectively, while the second terms represent the "loss" or decay of the amplitudes into states with one more particle. Because of the initial conditions $C_1(\eta_0) = 1; C_{n\neq 1}(\eta_0) = 0$, it follows that $d|C_2|^2/d\eta \propto \lambda^2 + \lambda^3 + \cdots$; $d|C_3|^2/d\eta \propto \lambda^4 + \lambda^5 + \cdots$ so that the (conformal) time dependence of the coefficients also follows a hierarchy: the three particle state "fills up" on time scales $\propto 1/\lambda^2$ larger than the two particle state, the four particle state on time scales $\propto 1/\lambda^2$ larger than the three particle state, etc.

Let us consider truncating the hierarchy beyond the three particles intermediate state, namely set $C_{[4]} = C_{[5]} = \cdots = 0$ along with all the other higher terms in the hierarchy. We then proceed to solve the equations from the bottom up with the initial conditions $C_1(\eta_0) = 1; C_{[2]}(\eta_0) = C_{[3]}(\eta_0) = \cdots = 0$. We obtain

$$C_3(\eta) = -i \int_{\eta_0}^{\eta} M_{32}(\eta') C_2(\eta')\, d\eta' \tag{5.6.6}$$

$$\dot{C}_2(\eta) = -iM_{21}(\eta) C_1(\eta) - \int_{\eta_0}^{\eta} d\eta_1 \sum_{[3]} M_{23}(\eta) M_{32}(\eta_1)\, C_2(\eta_1)\,. \tag{5.6.7}$$

The first term in (5.6.7) describes the build-up of the two-particle amplitude from the decay of the initial single particle state, whereas the second term describes the decay of the two-particle state into three particles via the cascade decay. Since the matrix elements are $\propto \lambda$ we can solve eqn. (5.6.7) iteratively in perturbation theory up to the order considered in the hierarchy, in order to understand the time scales,

$$\begin{aligned} C_2(\eta) &= -i \int_{\eta_0}^{\eta} M_{21}(\eta_1) C_1(\eta_1)\, d\eta_1 \\ &+i \int_{\eta_0}^{\eta} d\eta_1 \int_{\eta_0}^{\eta_1} d\eta_2 \int_{\eta_0}^{\eta_2} d\eta_3 \sum_{[3]} M_{23}(\eta_1) M_{32}(\eta_2) M_{21}(\eta_3)\, C_1(\eta_3) + \cdots\,. \end{aligned} \tag{5.6.8}$$

To make the arguments clear, let us consider Minkowski space time and early time scales so that $C_1 \simeq 1$. Then the two particle amplitude builds up $\propto \lambda t$ (with rate $\propto \lambda$), and from eqn. (5.6.6) we see that the three particle state builds up $\propto \lambda^2 t^2 \ll \lambda t$, clearly reflecting

that the population of the three particle state builds up much slower than that of the two particle state etc.

The build-up and decay integrals feature secular growth as $\eta \to 0$ (long cosmic time), and the second step in the Wigner-Weisskopf method provides a non-perturbative resummation of these processes: writing (5.6.7) as an integro-differential equation

$$\dot{C}_2(\eta) + \int_{\eta_0}^{\eta} \Sigma_{(2)}(\eta,\eta_1) C_2(\eta_1) d\eta_1 = -iM_{21}(\eta) C_1(\eta) \quad ; \quad \Sigma_{(2)}(\eta,\eta_1) = \sum_{[3]} M_{23}(\eta) M_{32}(\eta_1)\,, \tag{5.6.9}$$

and introducing the Markovian approximation as in eqn. (5.2.37-5.2.39) (the second approximation in the Wigner-Weisskopf method) we find

$$C_2(\eta) = -ie^{-\gamma_2(\eta)} \int_{\eta_0}^{\eta} M_{21}(\eta_1) C_1(\eta_1) e^{\gamma_2(\eta_1)}\, d\eta_1 \quad ; \quad \gamma_2(\eta) = \int_{\eta_0}^{\eta} \Sigma_{(2)}(\eta,\eta') d\eta'\,. \tag{5.6.10}$$

This compact expression reveals at once the build-up of the amplitude from C_1 and the eventual decay of the two-particle state encoded in $\gamma_2(\eta)$.

A simple perturbative expansion of this expression up to $\mathcal{O}(\lambda^4)$ reproduces (5.6.8) consistently with the Markovian approximation.

The last step is to insert this solution into (5.6.3), solve the integro-differential equation for C_1 and insert this solution into (5.6.8) and (5.6.6) respectively. Obviously this procedure leads to a very complicated expression that is not very illuminating. However progress can be made by introducing the perturbative solution (5.6.8), leading to the following integro-differential equation for C_1:

$$\begin{aligned} \dot{C}_1(\eta) &= -\int_{\eta_0}^{\eta} \sum_{[2]} M_{12}(\eta) M_{21}(\eta_1) C_1(\eta_1) \\ &+ \int_{\eta_0}^{\eta} d\eta_1 \int_{\eta_0}^{\eta_1} d\eta_2 \int_{\eta_0}^{\eta_2} d\eta_3 \sum_{[2],[3]} M_{12}(\eta) M_{23}(\eta_1) M_{32}(\eta_2) M_{21}(\eta_3)\, C_1(\eta_3) \end{aligned} \tag{5.6.11}$$

The first and second terms have a simple interpretation in terms of one and two loop self-energies as depicted in fig. (27) (only one two loop contribution is shown).

The dashed lines cut through multiparticle states and indicate similar rules to the Cutkosky rules of quantum field theory that relate the absorptive parts of self-energy diagrams to intermediate multiparticle states.

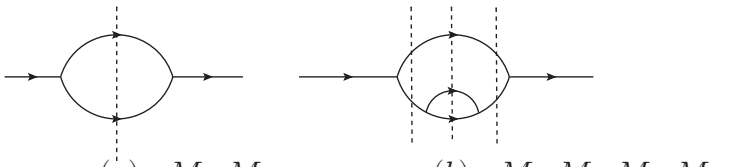


Figure 27: The contributions to C_1 showing the one and two loop contributions to the self-energy. The dashed lines represent intermediate states with two or three particles, corresponding to the matrix elements $M_{12}; M_{21}$ in (a), and similarly for (b). There are other two loop diagrams not shown.

In order to make progress in the solution of (5.6.11) the second part of the Wigner-Weisskopf method invokes a Markovian approximation, just as that described in section (5.2.2) implemented to lowest order. This approximation is again justified in a weak coupling expansion and is the statement that η derivatives of the coefficients are "slow" and can be systematically expanded perturbatively. The procedure follows the steps described by eqns. (5.2.37-5.2.40), integrating by parts the kernels of the integrals and keeping consistently up to $\mathcal{O}(\lambda^4)$ we find

$$\dot{C}_1(\eta) + W(\eta)C_1(\eta) = 0 \quad ; \quad C_1(\eta_0) = 1\,, \tag{5.6.12}$$

where

$$\begin{aligned} W(\eta) &= \int_{\eta_0}^{\eta} \sum_{[2]} M_{12}(\eta)M_{21}(\eta_1)d\eta_1 \Big[1 + \int_{\eta_0}^{\eta}\int_{\eta_0}^{\eta_1} \sum_{[2]} M_{12}(\eta)M_{21}(\eta_2)d\eta_1 d\eta_2\Big] \\ &+ \int_{\eta_0}^{\eta} d\eta_1 \int_{\eta_0}^{\eta_1} d\eta_2 \int_{\eta_0}^{\eta_2} d\eta_3 \sum_{[2],[3]} M_{12}(\eta)M_{23}(\eta_1)M_{32}(\eta_2)M_{21}(\eta_3)\,. \end{aligned} \tag{5.6.13}$$

The second term in the bracket in the first line arises from the derivative expansion of the term with the one-loop self-energy (see eqn. (5.2.39)); in ref.[175] this term is identified as a contribution to wave function renormalization. Therefore

$$C_1(\eta) = e^{-\int_{\eta_0}^{\eta} W(\eta')d\eta'}\,. \tag{5.6.14}$$

This expression provides a non-perturbative resummation of self-energies in real time up to two loops and includes the decay of the initial state into intermediate states with two and three particles.

In Minkowski space time, the initial state decays as $\propto e^{-\Gamma t}$ with $\Gamma = \lambda^2\gamma_2 + \lambda^4\gamma_3 + \cdots$ corresponding to the contribution to the self energy from the two particle intermediate states (one loop) three particle intermediate states (two loops) etc, highlighting that the probability of production of the two particle intermediate state occurs on a time scale $\propto 1/\lambda^2$, that of the three particle intermediate state on $\propto 1/\lambda^4$ etc.

Clearly the decay into two particle states occurs on shorter time scales as this process corresponds to the one-loop diagram, whereas decay into three particles occurs on much slower scales at this process corresponds to the two-loop contributions.

It remains to insert this solution into (5.6.8) and in turn insert the solution for $C_{[2]}$,into (5.6.6). Because the matrix elements $M_{ij} \propto \lambda$ it follows that if we take $C_1 \propto \mathcal{O}(\lambda^0)$, then $C_{[2]} \propto \lambda$; $C_{[3]} \propto \lambda^2 \cdots$. The quantum state obtained from the decay of a quanta with momentum $\vec{k}$ is given by

$$|\Psi(\vec{k},\eta)\rangle = C_1(\eta)|1_{\vec{k}}\rangle + \sum_{[2]} C_{[2]}(\eta)|[2]\rangle + \sum_{[3]} C_{[3]}(\eta)|[3]\rangle + \cdots , \tag{5.6.15}$$

The states $|[2]\rangle = |1_{\vec{p}}; 1_{\vec{k}-\vec{p}}\rangle$ and $|[3]\rangle = |1_{\vec{p}_1}; 1_{\vec{p}_2} : 1_{\vec{k}-\vec{p}_1-\vec{p}_2}\rangle$ and the sums over $[2],[3]$ are over $\vec{p}$ and $\vec{p}_1, \vec{p}_2$ respectively.

Thus the probability of a given two particle state is given by $|C_2(\eta)|^2 \propto \lambda^2$, of a given three particle state is $|C_3(\eta)|^2 \propto \lambda^4$, etc. This is exactly as in the case of multiphoton processes in quantum electrodynamics or of an atomic cascade of a multilevel atom. Each photon in the final state is associated a probability that is proportional to α_{em}, so multiphoton processes are suppressed by powers of the fine structure constant. In this case multiparticle final states are suppressed by powers of λ^2 for each extra particle in the final state. This is also the case in atmospheric air showers where very energetic particles decay via a cascade process where each branch of the cascade is down by a power of the coupling to the respective channel.

In the case of cascade decay in Minkowski space time, the probability of finding particles from a particular decay channel is given by the branching ratio of such channel Γ_c/Γ_{tot},

namely ratios of different powers of the couplings. Our result obviously entails the same physics: the probability of a state with three quanta is suppressed by λ^2 with respect to that with only two quanta, etc.

Furthermore, the explicit form of $W(\eta)$ (5.6.13) clearly shows the separation of time scales: the decay into two particles involves time scales $\propto 1/\lambda^2$ and is determined by the product of matrix elements $M_{21}M_{12}$ whereas the time scales for decay into three particle states is determined by the last term in (5.6.13) which implies time scales $\propto 1/\lambda^4$. Therefore there is a hierarchy both in the probability of multiparticle states and the time scales associated with their production from the decay of the parent particle. The cascade decay processes are controlled by the small coupling λ.

The entanglement entropy can now be calculated by obtaining the reduced density matrix by tracing over the superhubble degrees of freedom in the pure state density matrix $|\Psi(\vec{k},\eta)\rangle\langle\Psi(\vec{k},\eta)|$ and is a straightforward implementation of the steps described in the previous section with the technical complication of the integration over the super Hubble subset of momenta in the multiparticle contributions. This is only a technical difficulty but not a conceptual roadblock, since the contribution to the entanglement entropy from higher multiplicity states will be suppressed by high powers of the coupling λ. An illustrative example in Minkowski space time is the cascade decay $\pi^- \to \mu^- \overline{\nu}_\mu$; $\mu^- \to e^- \overline{\nu}_e \nu_\mu$: whereas the pion decays on a time scale $\simeq 2.8 \times 10^{-8}$ secs the muon decays on a time scale $\simeq 2.2 \times 10^{-6}$ secs therefore during a long time interval 10^{-8} secs $\leq t \leq 10^{-6}$ secs the two particle state $|\mu^-, \overline{\nu}_\mu\rangle$ yields the largest contribution to the quantum state.

Furthermore, the unitarity relation (5.2.48) entails that

$$|C_1(\eta)|^2 + \sum_{[2]} |C_{[2]}(\eta)|^2 + \sum_{[3]} |C_{[3]}(\eta)|^2 + \cdots = 1\,, \tag{5.6.16}$$

which was confirmed in the previous section to leading order in the coupling and Δ.

In summary: the cascade decay is controlled by the perturbative nature of the interaction, the probability for multiparticle states being suppressed by powers of the coupling constant and the time scales associated with the formation of multiparticle states widely separated by larger powers of $1/\lambda$. Furthermore, in the case under consideration here, the physical

momentum of the initial state is taken to remain deep inside the Hubble radius at all times during inflation. At any large but fixed (conformal time) the initial state maintains a small but non-vanishing population, a two particle state being populated with probability λ^2, a given three particle state with probability $\propto \lambda^4$ etc. Therefore if the quasi-de Sitter inflationary stage lasts a finite (say $\simeq 60$) number of e-folds, the quantum state will be a linear superposition of many particle states and unitarity implies that each state features a perturbatively small population. An interesting and conceptually puzzling situation arises in the case of eternal de Sitter, since in this case, at asymptotically long times all states would have decayed to vanishing probability in clear contradiction with unitarity, but in this case all physical momenta eventually also become superHubble. Perhaps this puzzling aspect is related to the intriguing results of ref.[187] and deserves to be studied further.

While we have established a roadmap and a "proof of principle" of the method, undoubtedly there are several aspects that merit a deeper study such as infrared enhancement from superhorizon modes, the issue of unitarity in eternal de Sitter, the detailed aspects of the (conformal) time dependence of the amplitudes of multiparticle states etc. We postpone the study of these more technical details of the higher order processes to a future article.

5.7 DISCUSSION AND FURTHER QUESTIONS

Possible relation to Non-Gaussianity.

The cubic interaction vertex suggests a relation between the decay amplitude (see fig (24)) and the non-gaussian bispectrum which is the three point function of the field. The relationship with the bispectrum becomes more clear by introducing $G(k,\eta,\eta') = g^*_\nu(k;\eta)g_\nu(k;\eta')$ from which it follows that

$$\int_{\eta_0}^{\eta} \Sigma(k,\eta,\eta')\, d\eta' \propto \int \frac{d^3p}{(2\pi)^3} \int \frac{d\eta'}{H\eta'}\, G(k,\eta,\eta')\, G(p,\eta,\eta')\, G(|\vec{k}-\vec{p}|,\eta,\eta')\,. \tag{5.7.1}$$

The imaginary part of the η'-integral is proportional to the *bispectrum* of the scalar field[206, 207]. The main difference is that the self-energy is the integral over one of the momenta.

In particular the leading order in Δ, namely the contribution from the infrared enhanced, superhorizon modes, is determined by the highly squeezed limit shown in fig. (28), which corresponds to the local limit of the non-gaussian correlator.

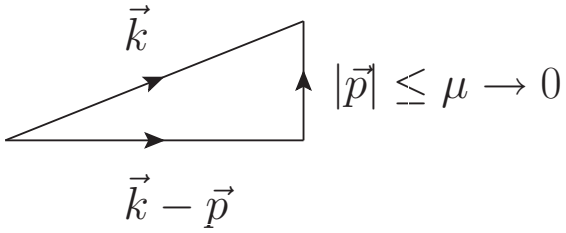


Figure 28: Triangle of momenta for the bispectrum (see eqn.(5.7.1)) integration over $|\vec{p}| < \mu \to 0$ corresponds to the highly squeezed limit and yields the pole in Δ.

This connection highlights that this local limit is describing *correlations* between subhorizon and superhorizon modes, these are the correlations that yield the entanglement entropy upon tracing over the superhorizon degrees of freedom.

There are important differences between the scalar field theory with cubic interaction studied here, and the cubic interactions of curvature perturbations in the theory of non-gaussian fluctuations[206, 207], the main difference being both spatial and (conformal) time derivatives in the interactions. However, the study of ref.[208] found that transition probabilities of curvature perturbations (in single field slow roll inflation) are suppressed by slow roll parameters but enhanced by infrared logarithms, which are similar to those emerging in the $\Delta \to 0$ limit in our study (corresponding to massless fluctuations), thus suggesting that the results obtained above *may* apply to the decay of curvature perturbations and superhorizon entanglement and concomitant entanglement entropy.

Is the information retrieved upon horizon entry?:

An important feature of inflationary cosmology is that physical wavelengths that cross the Hubble radius during inflation, re-enter the Hubble radius (now the particle horizon) during radiation or matter domination and these quantum fluctuations are the seeds of temperature anisotropies and inhomogeneities.

The entanglement entropy that we have studied is a measure of the correlations be-

tween the entangled subhorizon and superhorizon degrees of freedom as a consequence of interactions, which brings the question of whether upon re-entry the fluctuation modes that were superhorizon during inflation "bring back" the quantum correlations and if so how are these manifest in the power spectrum of the CMB? Furthermore, going from quantum fluctuations of the curvature (or gravitational potential) to temperature fluctuations entails replacing quantum averages by statistical averages. Thus it is a relevant question whether this statistical averaging includes the quantum correlations from entanglement. Last but not least, if the quantum states can decay, it is conceivable that the lack of power in the low multipoles which is present in the cosmological data and has been persistent in the statistical analysis of WMAP7[209], WMAP9[210] and Planck[63] which reports a power deficit at low multipole with $2.5-3\,\sigma$ significance and a recent statistical analysis of the combined dataset[211], may be due to the decay of the quantum fluctuations during the inflationary stage. Our study applies to a scalar field in de Sitter space time and in order to answer this question the analysis presented here must be applied to the case of scalar perturbations.

Intensity correlations?:) Furthermore, and in relation with the question above, it is a tantalizing possibility that the superhorizon correlations become manifest as intensity correlations leading to interference phenomena akin to the Hanbury-Brown-Twiss effect discussed in ref.[195]. If this is the case what would be the observable consequences of the correlations between sub and superhorizon degrees of freedom.

An infinite cascade?: The discussion of the cascade process offered above is mainly based on the physical aspects of cascade decays in Minkowski space time (for example shower cascade). However, infrared effects may modify this picture justifying a deeper understanding of how the infrared enhancements modify the higher order multiparticle processes in the cascade decay. Furthermore an interesting remaining question is how unitarity is manifest in the (formal) case of eternal de Sitter inflation in which the cascade processes would continue forever perhaps resulting in a quantum state of infinitely many particles but with infinitesimally small probabilities. This question clearly merits a continued effort to understand these aspects in view of the results of ref.[187].

5.8 CONCLUSIONS:

In inflationary cosmology all particle states *decay* as a consequence of the lack of a global time-like Killing vector which would in turn enforce kinematic thresholds. In this article we have studied the entanglement entropy from the decay of single particle states during de Sitter inflation in a theory of a light scalar field with $M^2 \ll H^2$ and cubic interactions. The quantum state that describes the single particle decay and the produced particles is a two-particle state entangled by momentum conservation. We have extended and generalized the Wigner-Weisskopf method used in the treatment of spontaneous decay of atomic states to the realm of quantum field theory in an expanding cosmology, and implemented this method to obtain the quantum state that describes the decay of the parent particle and the production of the daughter particles. We showed in detail that this non-perturbative approximation is *manifestly unitary*. The amplitudes for the two-particle entangled state features infrared enhancements that are manifest as poles in $\Delta = M^2/3H^2$ as a consequence of the emission of superhorizon quanta and we implement a consistent expansion in Δ to leading order to obtain the (pure state) density matrix that describes the decay of the parent and production of daughter particles. When the parent particle's wavelength is inside the horizon, the density matrix elements for the produced particles are dominated by the contribution of *superhorizon* momenta of one of the daughter particles, describing entanglement, correlation and coherences across the horizon. Tracing the pure state density matrix over the superhorizon modes we obtain a *mixed state* density matrix from which we calculate the Von Neumann entanglement entropy, which describes the loss of information from the correlations between sub and superhorizon modes due to the non-observation of these latter states. We find that the entanglement entropy is enhanced in the infrared by a factor of $\ln[1/\Delta]$ and grows logarithmically with the physical volume as a consequence of more modes crossing the Hubble radius during the inflationary stage.

The generalization to the description of single particle states in terms of wavepackets spatially localized within the Hubble radius but localized in momentum was provided. Under the conditions that the average wavevector of the wave packet be associated with subHubble wavelengths all throughout the near de Sitter stage, we showed the equivalence between the

plane wave and wave packet description and assessed the corrections in terms of the ratio of the width of the wavepacket in momentum space and the average momentum associated with the single particle state.

The lack of kinematic thresholds implies that particle decay occurs in a cascade process, namely $1 \to 2 \to 3 \cdots$. We have extended the Wigner-Weisskopf method to establish a framework to study the cascade decay and analyzed in detail the process up to a three particle branching in the cascade, but the results are quite general. We showed that for weak coupling (here we considered a cubic coupling) the probability of multiparticle states is suppressed by powers of the coupling, for example in the case of cubic coupling the three particle state is suppressed by $\mathcal{O}(\lambda^2)$ with respect to the two particle state, the four particle $\mathcal{O}(\lambda^4)$ etc. We have established a relation between the different multiparticle processes and higher order loop contributions in the self-energy, just as in the case of Cutkosky rules in Minkowski space-time. This relation clearly shows that just as the probability of higher multiparticle states is suppressed by high powers of the coupling, the time scales for decay into higher multiplicity states are widely separated by inverse powers of λ^2. Therefore the cascade decay is controlled by the weak coupling, just as multiphoton processes in QED, however important questions raised above remain, justifying continued study of these issues.

This study of the *superhorizon entanglement entropy* from particle decay bridges two concepts previously explored in the literature: the entanglement between spatially separated but correlated regions, in our case the correlations between sub and superhorizon quanta of the daughter particles, akin to the superhorizon correlations studied in ref.[172], and the momentum-space entanglement studied in ref.[173, 198]. In our study the entanglement entropy is a result of *both* types of concepts, linked together by the interactions but with the distinct aspect of being a non-equilibrium process as a consequence of the cosmological expansion.

While at this stage we do not *yet* see a clear observational consequence of the entanglement entropy beyond the theoretical conceptual aspect of information loss from the correlations and superhorizon entanglement, the exploration of potential observational consequences along with the questions raised above are worthy of further and deeper study, on which we

expect to report in the future.

6.0 TIME EVOLUTION OF CASCADE DECAY.

Based on: (ref. [431])

D. Boyanovsky, L. Lello, New Journal of Physics 16 (2014) 063050

6.1 INTRODUCTION

The decay of unstable or metastable states via a cascade $A \to BC \to BXY$ is of interdisciplinary importance in particle physics, quantum optics and cosmology. Early studies in particle physics proposed to use the time evolution of intermediate (resonant) states in cascade decays of heavy (B) mesons to study aspects of CP violation, mixing phenomena[213, 214, 215, 216, 217, 218] and CPT violation[219, 220]. The Belle collaboration[156] has reported on remarkably precise measurements of (EPR) entanglement and correlations in cascade decays $\Upsilon(4S) \to B^0\overline{B}^0 \to (l, J/\Psi, K)$ via the analysis of the time dependence of the flavor asymmetry[221]. The BaBar collaboration[157, 158] has reported on the first direct measurement of time reversal violation in the $B^0\overline{B}^0$ system from $\Upsilon(4S)$ decay at rest by studying the time dependent correlations between the members of the entangled $B^0\overline{B}^0$ pairs also in a cascade decay.

More recently cascade decays have been proposed as possible mechanisms of CP and lepton flavor violation mediated by heavy sterile neutrinos as resonant states[222] and as possible solutions to the LSND/MiniBooNe anomalies[124]. Cascade decays may probe new particles beyond the standard model via kinematic edges associated with the new degrees of

freedom in the intermediate states[223].

In quantum optics and cavity Quantum Electrodynamics, the cascade decay of multi-level atoms or quantum dot systems is studied as a source of correlated photons and entanglement [204, 224, 212, 225, 226, 227] with potential applications in quantum information. The possibility of observation of polarization entangled photon pairs in cascade decay in quantum dots has recently been advanced, such a measurement is, fundamentally, similar to the observation of CP violating amplitudes in the time evolution of a cascade decay in B-mesons and similar meson systems addressed in refs.[213]-[218].

In inflationary cosmology the rapid expansion of the Universe entails that there is no time-like Killing vector and as a consequence of the lack of kinematic thresholds quanta of a field can decay into quanta of the *same* field [228, 229, 175, 194, 192] leading to the recent suggestion of cascade decay of inflaton quanta during inflation[311] and the concomitant kinematic entanglement of the produced particles.

Motivation and Goals: In the analysis of cascade decays in refs.[213, 214, 215, 216, 217], the time evolution of the cascade is analyzed as a *sequential* series of events. Consider an initial state I decaying into an intermediate state M, which in turn decays into a final state f, the amplitude for such process is proposed to be

$$A[I \stackrel{t_I}{\to} M \stackrel{t_M}{\to} f] = e^{-iW_I\, t_I}\; e^{-iW_M\, t_M}\; \mathcal{M}_{I\to M}\; \mathcal{M}_{M\to f} \tag{6.1.1}$$

where t_I is the time at which the initial state I decays into the intermediate resonant state M and t_M is the time at which the intermediate state M decays into the final state in their respective rest frames, $W_{I,M} = m_{I,M} - i\Gamma_{I,M}$ are the complex energies of the corresponding states and $\mathcal{M}$ are the corresponding transition amplitudes (we neglect here the possibility of mixing in the initial, intermediate or final states).

While this sequential characterization may be phenomenologically suitable to the description of experimental situations in which there is a wide separation in time scales as a consequence of large differences in widths and masses, it clearly *assumes* that the decays occur at specific (proper) times t_I, t_M in sequence.

However, in a general case, this is at odds with the description of radiative cascades in multi-level systems in quantum optics[204, 224, 212] and with the description of decay of an unstable (or metastable) state as a continuous process with a (generally exponential) time distribution in which the amplitude for the parent particle decays exponentially and the amplitude of the daughter (resonant state) increases continuously on a similar time scale.

The interdisciplinary relevance of cascade decay motivates us to study the time evolution within a framework that is suitable to extension to the realm of cosmology and that could prove useful in other areas such as quantum optics and condensed matter physics.

The goal of this article is to study the full time evolution of a cascade decay process directly from the quantum mechanical evolution of an initial quantum state. In particular we address the important issue of unitarity in the time evolution from the initial to the final state, focusing on the time evolution of the amplitudes and populations of the intermediate resonant and final states and how unitarity is manifest in these amplitudes.

For this purpose we provide a non-perturbative quantum field theoretical generalization of the methods ubiquitous in quantum optics[204, 224, 212, 225] adapted to the realm of a full quantum field theory. Furthermore, one of our objectives is to provide a non-perturbative method to study the time evolution directly in real time that would be suitable for applications in cosmology where the expansion of the Universe introduces an explicit time dependence in the evolution Hamiltonian[229, 192, 311].

In this article we consider the case of only one resonant intermediate state to introduce and develop the methods and to exhibit the main physical processes in a simpler setting, postponing the study of mixing and oscillations in the resonant state for future study.

Brief summary of results: We implement a non-perturbative method that yields the full time evolution of an initial state of a decaying parent particle in a model of generic fields that incorporates the main features of a cascade decay via a resonant state $\pi \to \phi_1\phi_2 \to \phi_2\chi_1\chi_2$ although the results are general. We obtain the time evolution of the amplitudes for the intermediate and final states and show explicitly that the method is akin to a Dyson-type resummation of self-energies and fulfills unitarity. Unitary time evolution is manifest in a transfer or "flow" of probability or population from the initial through the intermediate

resonant to the final state. We analyze in detail the evolution of the populations on the different time scales. The population of the resonant state grows at early times reaching a maximum at $t^* = \ln[\Gamma_\pi/\Gamma_{\phi_1}]/(\Gamma_\pi - \Gamma_{\phi_1})$ and decaying to the final state.

The build-up of the population of the intermediate resonant state depends crucially on the ratio of decay widths $\Gamma_\pi/\Gamma_{\phi_1}$. For $\Gamma_\pi/\Gamma_{\phi_1} \gg 1$ there is a "bottleneck" in the sense that the population of the resonant state builds up to nearly saturate unitarity on a time scale $t \simeq t^*$ and decays into the final state on much longer time scales $\simeq 1/\Gamma_{\phi_1}$. In the opposite limit, the population of the initial state is transferred almost directly to the final state with a very small build-up of the population of the intermediate resonant state.

The final asymptotic state after both the parent particle and intermediate resonant state decays is a many particle state featuring quantum entanglement and correlations among the final particles.

We also provide a quantum field theoretical generalization of the Wigner-Weisskopf method that provides a similar type of non-perturbative resummation directly in real time and yields the same results but is amenable of implementation in cosmology where the interaction Hamiltonian is explicitly dependent on time.

We conjecture on possible phenomenological consequences in particular for heavy sterile neutrinos as intermediate resonant states in pseudoscalar decays.

6.2 THE MODEL

We consider a model of generic real scalar fields $\pi, \phi_{1,2}, \chi_{1,2}$ of masses $M_\pi; m^\phi_{1,2}, m^\chi_{1,2}$ respectively to study the relevant phenomena in the simplest setting and to focus on the main aspects of the method and physical processes, however the main results will be argued to be general.

The total Hamiltonian is $H = H_0 + H_I$ with H_0 the free field Hamiltonian and

$$H_I = \int d^3x \Big\{ g_\pi \, \pi(x)\phi_1(x)\phi_2(x) + g_\phi \, \phi_1(x)\chi_1(x)\chi_2(x) \Big\} \, . \tag{6.2.1}$$

Let us consider an initial state with one π particle of momentum $\vec{k}$ and the vacuum for the other fields, namely

$$\big|\Psi(t=0)\rangle = \big|\pi_{\vec{k}}\rangle\,. \tag{6.2.2}$$

Upon time evolution this state evolves into $\big|\Psi(t)\rangle$ obeying

$$\frac{d}{dt}\big|\Psi(t)\rangle = -i(H_0+H_I)\big|\Psi(t)\rangle\,, \tag{6.2.3}$$

when $m_\pi > m_{\phi_1}+m_{\phi_2}\ ;\ m_{\phi_1} > m_{\chi_1}+m_{\chi_2}$ the interaction Hamiltonian (6.2.1) describes the cascade process depicted in fig.29.

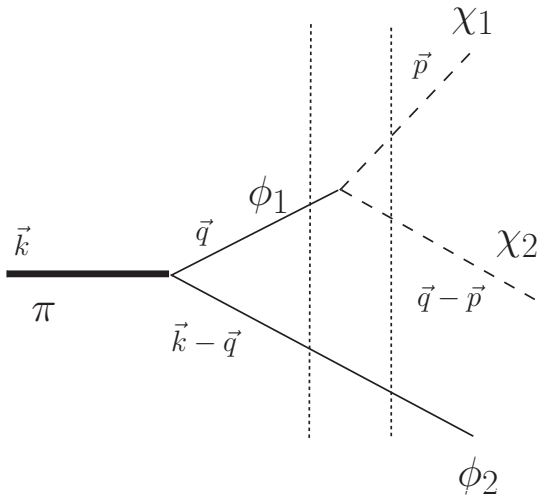


Figure 29: Cascade decay $\pi \to \phi_1\phi_2; \phi_1 \to \chi_1\chi_2$. The dashed lines depict the intermediate two particle state and the final three particle state.

At any given time the state $\big|\Psi(t)\rangle$ can be expanded in Fock states of the non-interacting Hamiltonian H_0

$$\big|\Psi(t)\rangle = \sum_n C_n(t)\,\big|n\rangle \quad ; \quad H_0\big|n\rangle = E_n\big|n\rangle \tag{6.2.4}$$

the time evolution of the coefficients $C_n(t)$ is obtained from (6.2.3), and projecting onto the states $\big|n\rangle$ namely

$$\dot{C}_n(t) = -iE_n\,C_n(t) + \sum_\kappa \langle n\big|H_I\big|\kappa\rangle\,C_\kappa(t)\,. \tag{6.2.5}$$

This is an infinite hierarchy of equations, progress is made by truncating the hierarchy at a given order and solving the coupled equations, this method is equivalent to a *Dyson resummation of self-energy diagrams* as will be seen clearly and in detail below.

The time evolution of the initial state depicted up to second order in the interaction in fig.29 leads to considering the following form

$$\big|\Psi(t)\rangle = C_\pi(\vec{k},t)\big|\pi_{\vec{k}}\rangle + \sum_{\vec{q}} C_{\phi\phi}(\vec{k},\vec{q};t)\,\big|\phi_{1,\vec{q}}\phi_{2,\vec{k}-\vec{q}}\rangle + \sum_{\vec{q};\vec{p}} C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t)\,\big|\phi_{2,\vec{k}-\vec{q}}\chi_{1,\vec{p}}\chi_{2,\vec{q}-\vec{p}}\rangle + \cdots \tag{6.2.6}$$

where the dots stand for many particle states that emerge in higher order in H_I. The matrix elements of the interaction Hamiltonian in (6.2.5) describe transitions between single and multiparticle states, and the functions $C_\pi(\vec{k},t); C_{\phi\phi}(\vec{k},\vec{q};t); C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t)$ represent the amplitudes of the initial single particle state and multiparticle states in the time evolved state. The initial conditions on these amplitudes are

$$C_\pi(\vec{k},0) = 1\ ;\ C_{\phi\phi}(\vec{k},\vec{q};0) = 0\ ;\ C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};0) = 0\,. \tag{6.2.7}$$

The evolution equations for the amplitudes are obtained by projection, as in (6.2.5). Introducing the shorthand definitions

$$M_{\pi\to\phi_1\phi_2}(\vec{k},\vec{q}) \equiv \langle\pi_k|H_I|\phi_{1,\vec{q}}\,\phi_{2,\vec{k}-\vec{q}}\rangle \tag{6.2.8}$$

$$M_{\phi_1\to\chi_1\chi_2}(\vec{q},\vec{p}) \equiv \langle\phi_{1,\vec{q}}|H_I|\chi_{1,\vec{p}}\,\chi_{2,\vec{q}-\vec{p}}\rangle \tag{6.2.9}$$

and

$$E_{\phi\phi}(\vec{k},\vec{q}) \equiv E^{\phi_1}_{\vec{q}} + E^{\phi_2}_{\vec{k}-\vec{q}} \tag{6.2.10}$$

$$E_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p}) \equiv E^{\phi_2}_{\vec{k}-\vec{q}} + E^{\chi_1}_{\vec{p}} + E^{\chi_2}_{\vec{q}-\vec{p}} \tag{6.2.11}$$

where $E_{\vec{k}}$ are the single particle energies for π,ϕ,χ respectively we find

$$\dot{C}_\pi(\vec{k},t) = -iE^\pi_k\,C_\pi(\vec{k},t) - i\sum_{\vec{q}} M_{\pi\to\phi_1\phi_2}(\vec{k},\vec{q})\,C_{\phi\phi}(\vec{k},\vec{q};t) \tag{6.2.12}$$

$$\dot{C}_{\phi\phi}(\vec{k},\vec{q};t) = -iE_{\phi\phi}(\vec{k},\vec{q})\,C_{\phi\phi}(\vec{k},\vec{q};t) - iM^{*}_{\pi\to\phi_1\phi_2}(\vec{k},\vec{q})\,C_{\pi}(\vec{k},t) - i\sum_{\vec{p}} M_{\phi_1\to\chi_1\chi_2}(\vec{q},\vec{p})\,C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t) \tag{6.2.13}$$

$$\dot{C}_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t) = -iE_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p})\,C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t) - iM^{*}_{\phi_1\to\chi_1\chi_2}(\vec{q},\vec{p})\,C_{\phi\phi}(\vec{k},\vec{q};t) \tag{6.2.14}$$

We have truncated the hierarchy of equations up to second order in the interaction Hamiltonian (6.2.1), thereby neglecting the higher order branches with four particles etc. As it will be demonstrated below, the solution of the coupled hierarchy of equations up to this order provides Dyson resummations of *both* the propagator of the π particle *and* the propagator for the ϕ_1 particle which corresponds to the resonant intermediate state in the cascade decay process. This will become clear below.

In what follows we will suppress the arguments of the various functions introduced above to simplify notation.

It proves convenient to solve these coupled set of equations by Laplace transform as befits an initial value problem. Defining the Laplace transform of $C(t)$ as

$$\tilde{C}(s) = \int_0^\infty e^{-st} C(t) dt \tag{6.2.15}$$

and with the initial conditions (6.2.7) the set of coupled equations (6.2.12-6.2.14) becomes

$$\tilde{C}_{\pi}(\vec{k},s)\,[s+iE^{\pi}_{k}] = 1 - i\sum_{\vec{q}} M_{\pi\to\phi_1\phi_2}(\vec{k},\vec{q})\,\tilde{C}_{\phi\phi}(\vec{k},\vec{q};s)\,, \tag{6.2.16}$$

$$\tilde{C}_{\phi\phi}(\vec{k},\vec{q};s)[s+iE_{\phi\phi}] = -iM^{*}_{\pi\to\phi_1\phi_2}(\vec{k},\vec{q})\,\tilde{C}_{\pi}(\vec{k};s) - i\sum_{\vec{p}} M_{\phi_1\to\chi_1\chi_2}(\vec{q},\vec{p})\,\tilde{C}_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};s)\,, \tag{6.2.17}$$

$$\tilde{C}_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};s)[s+iE_{\phi\chi\chi}] = -iM^{*}_{\phi_1\to\chi_1\chi_2}(\vec{q},\vec{p})\,\tilde{C}_{\phi\phi}(\vec{k},\vec{q};s)\,. \tag{6.2.18}$$

We now solve this system of algebraic equations from the bottom up obtaining,

$$\tilde{C}_{\pi}(\vec{k};s) = \frac{1}{s+iE^{\pi}_{k}+i\Sigma^{\pi}(\vec{k};s)}\,, \tag{6.2.19}$$

$$\tilde{C}_{\phi\phi}(\vec{k},\vec{q};s) = -i\frac{M^*_{\pi\to\phi_1\phi_2}\,\tilde{C}_{\pi}(\vec{k};s)}{s+iE_{\phi\phi}+i\Sigma^{\phi}(\vec{k};s)}\,, \tag{6.2.20}$$

$$\tilde{C}_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};s) = -i\frac{M^*_{\phi_1\to\chi_1\chi_2}\,\tilde{C}_{\phi\phi}(\vec{k};s)}{s+iE_{\phi\chi\chi}}\,, \tag{6.2.21}$$

where the self-energies are given by

$$i\Sigma^{\pi}(\vec{k};s) = \sum_{\vec{q}}\frac{|M_{\pi\to\phi_1\phi_2}|^2}{s+iE_{\phi\phi}+i\Sigma^{\phi}(\vec{k},\vec{q};s)}\,, \tag{6.2.22}$$

and

$$i\Sigma^{\phi}(\vec{k},\vec{q};s) = \sum_{\vec{p}}\frac{|M_{\phi_1\to\chi_1\chi_2}|^2}{s+iE_{\phi\chi\chi}}\,. \tag{6.2.23}$$

These equations have a familiar interpretation in terms of Dyson resummations of self-energy diagrams depicted in fig.(30): the irreducible self-energy for the π-field is in terms of the full propagator of the (resonant) ϕ_1 particle which is itself given by a Dyson resummation of irreducible self-energy diagrams involving a loop of the final state particles. It is at this stage that the equivalence between the truncation of the hierarchy of equations and the resummation in terms of a Dyson series becomes manifest. Therefore, the solution of the truncated hierarchy up to the given (second) order, provides a non-perturbative resummation of self-energy diagrams both for the decaying parent particle *and* the intermediate resonant state. The π self-energy (6.2.22) includes the self-energy correction to the (intermediate) ϕ_1 state.

6.2.1 Time evolution:

The time evolution is obtained by performing the inverse Laplace transform:

$$C(t) = \int_{\mathcal{C}}\frac{ds}{2\pi\,i}\;e^{st}\;\tilde{C}(s) \tag{6.2.24}$$

where $\mathcal{C}$ stands for the Bromwich contour parallel to the imaginary axis in the complex s plane and to the right of all the singularities of the function $\tilde{C}(s)$.

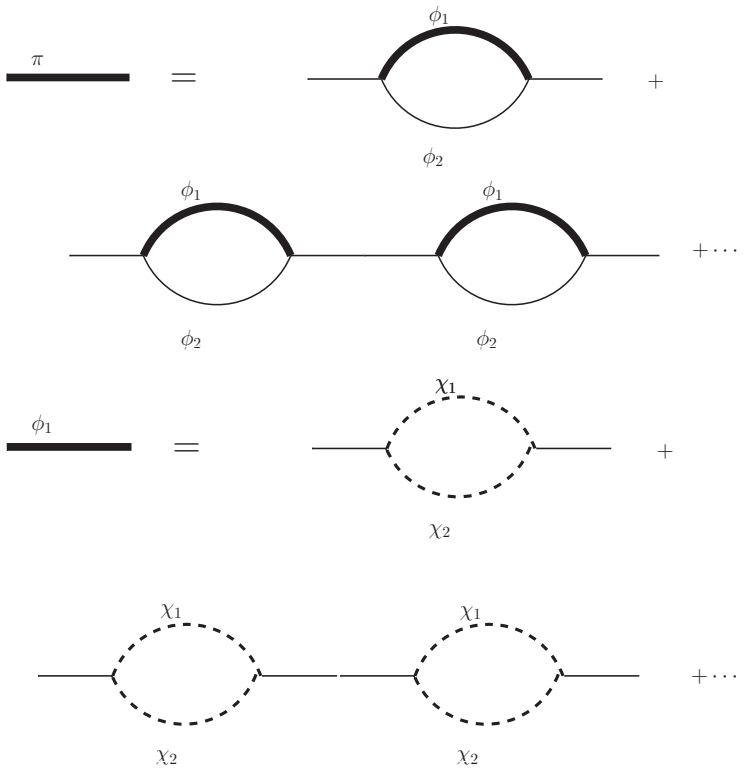


Figure 30: π propagator, the thick ϕ_1 line is the full ϕ_1 propagator with self-energy resummation.

With the purpose of using the convolution theorem

$$\int_{\mathcal{C}} \frac{ds}{2\pi\, i}\; e^{st}\; \tilde{F}(s)\, \tilde{G}(s) = \int_0^t F[t-t']\, G[t']dt' \tag{6.2.25}$$

where $F[t], G[t]$ are the inverse Laplace transforms of $\tilde{F}(s), \tilde{G}(s)$ respectively, it proves convenient to rewrite (6.2.20,6.2.21) as

$$\tilde{C}_{\phi\phi}(\vec{k},\vec{q};s) = -iM^*_{\pi\to\phi_1\phi_2}\, \tilde{C}_{\pi}(\vec{k};s)\; \tilde{G}_{\phi\phi}(s) \quad ; \quad \tilde{G}_{\phi\phi}(s) = \frac{1}{s + iE_{\phi\phi} + i\Sigma^{\phi}(s)} \tag{6.2.26}$$

$$\tilde{C}_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};s) = -iM^*_{\phi_1\to\chi_1\chi_2}\, \tilde{C}_{\phi\phi}(\vec{k};s)\, \tilde{G}_{\phi\chi\chi}(s) \quad ; \quad \tilde{G}_{\phi\chi\chi}(s) = \frac{1}{s + iE_{\phi\chi\chi}} \tag{6.2.27}$$

The amplitudes $\tilde{C}(s)$ given by eqns. (6.2.19-6.2.21) generally feature branch cuts above multiparticle thresholds and complex poles with Re$s \leq 0$ namely either along the imaginary axis or to its left as befits stable particles, multiparticle cuts (along the imaginary axis) or decaying resonances. Therefore the Bromwich contour corresponds to integration along the path parallel to the imaginary axis $s = i\omega + \varepsilon$; $-\infty \leq \omega \leq \infty$; $\varepsilon \to 0^+$ and

$$C(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, e^{i\omega t}\, \tilde{C}(s = i\omega + \varepsilon) \tag{6.2.28}$$

where the contour must now be closed in the upper half $\omega-$ plane.

The first step in obtaining the time evolution of the amplitudes is to obtain $C_\pi(t)$ for which we need to identify the singularities in

$$\tilde{C}_\pi(s = i\omega + \varepsilon) = \frac{1}{i\omega + iE_k^\pi + i\Sigma^\pi(\vec{k}, i\omega + \varepsilon) + \varepsilon}\,. \tag{6.2.29}$$

The self-energy features a two-particle cut and the π particle becomes a resonance if E_k^π is embedded in the two-particle continuum. For a weakly coupled theory this resonance is described by a complex pole in the upper half plane very near the real axis, since in absence of perturbations the pole is at $\omega = -E_k^\pi$. Consistently in perturbation theory we write $\omega = -E_k^\pi$ in the argument of Σ^π, with

$$i\Sigma^\pi(\vec{k}, -iE_k^\pi + \varepsilon) = \sum_{\vec{q}} \frac{|M_{\pi\to\phi_1\phi_2}|^2}{-iE_k^\pi + iE_{\phi\phi} + i\Sigma^\phi(\vec{k}, \vec{q}; -iE_k^\pi + \varepsilon) + \varepsilon} \tag{6.2.30}$$

and from (6.2.23) we find

$$i\Sigma^\phi(\vec{k}, \vec{q}; s = -iE_k^\pi + \varepsilon) = i\sum_{\vec{p}} \frac{|M_{\phi_1\to\chi_1\chi_2}|^2}{(E_k^\pi - E_{\phi\chi\chi}) + i\varepsilon} = i\delta E + \frac{1}{2}\,\gamma(\vec{k}, \vec{q}) \tag{6.2.31}$$

$$\delta E = \sum_{\vec{p}} \mathcal{P}\, \frac{|M_{\phi_1\to\chi_1\chi_2}|^2}{(E_k^\pi - E_{\phi\chi\chi})}\,, \tag{6.2.32}$$

$$\gamma(\vec{k}, \vec{q}) = 2\pi \sum_{\vec{p}} |M_{\phi_1\to\chi_1\chi_2}|^2\, \delta(E_k^\pi - E_{\phi\chi\chi})\,, \tag{6.2.33}$$

we note that δE is *not* the renormalization of the ϕ_1 energy (mass) and $\gamma(\vec{k}, \vec{q})$ is *not* its decay width, because in the denominator of (6.2.32) and the argument of the delta function

in (6.2.33) is $E^{\pi}_{k} - E^{\phi_2}_{\vec{k}-\vec{q}} - E^{\chi_1}_{\vec{p}} - E^{\chi_2}_{\vec{q}-\vec{p}}$ instead of being $E^{\phi_1}_{\vec{q}} - E^{\chi_1}_{\vec{p}} - E^{\chi_2}_{\vec{q}-\vec{p}}$ (see below for clarification on this point).

Inserting this expression into (6.2.30) we find

$$i\Sigma^{\pi}(\vec{k}, -iE^{\pi}_{k} + \varepsilon) = i\Delta E^{\pi} + \frac{\Gamma_{\pi}(k)}{2} \tag{6.2.34}$$

where the energy shift

$$\Delta E^{\pi} = \sum_{\vec{q}} \frac{|M_{\pi\to\phi_1\phi_2}|^2 \ (E^{\pi}_{k} - E_{\phi\phi} - \delta E)}{\left(E^{\pi}_{k} - E_{\phi\phi} - \delta E\right)^2 + \left(\frac{\gamma(\vec{k},\vec{q})}{2}\right)^2} \tag{6.2.35}$$

is absorbed into a renormalization of the π mass, and

$$\begin{aligned} \Gamma_{\pi}(k) &= \sum_{\vec{q}} \frac{|M_{\pi\to\phi_1\phi_2}|^2 \ \gamma(\vec{k},\vec{q})}{\left(E^{\pi}_{k} - E_{\phi\phi} - \delta E\right)^2 + \left(\frac{\gamma(\vec{k},\vec{q})}{2}\right)^2} \\ &= 2\pi \sum_{\vec{q}} \sum_{\vec{p}} \frac{|M_{\pi\to\phi_1\phi_2}|^2 \ |M_{\phi_1\to\chi_1\chi_2}|^2 \ \delta(E^{\pi}_{k} - E_{\phi\chi\chi})}{\left(E^{\pi}_{k} - E_{\phi\phi} - \delta E\right)^2 + \left(\frac{\gamma(\vec{k},\vec{q})}{2}\right)^2} \end{aligned} \tag{6.2.36}$$

is the decay width of the π particle.

As mentioned above $\gamma(\vec{k},\vec{q})$ is *not* the width of the ϕ_1 resonance, and (6.2.32) is *not* the renormalization of the ϕ_1 energy (mass renormalization), however in perturbation theory Γ_{π} (6.2.36) becomes resonant as $E^{\pi}_{k} \to E_{\phi\phi}$, and the sum is dominated by this resonance. Near this resonance one can replace $E^{\pi} \to E_{\phi\phi}$ into (6.2.32,6.2.33) and recognizing from (6.2.10,6.2.11) that $E_{\phi\phi} - E_{\phi\chi\chi} = E^{\phi_1}_{\vec{q}} - E^{\chi_1}_{\vec{p}} - E^{\chi_2}_{\vec{q}-\vec{p}}$ it follows that *near this resonance*

$$\delta E \to \Delta E^{\phi_1} = \sum_{\vec{p}} \mathcal{P} \ \frac{|M_{\phi_1\to\chi_1\chi_2}|^2}{(E^{\phi_1}_{\vec{q}} - E^{\chi_1}_{\vec{p}} - E^{\chi_2}_{\vec{q}-\vec{p}})} \tag{6.2.37}$$

is the energy shift absorbed into a renormalization of the ϕ_1 mass and

$$\gamma(\vec{k},\vec{q}) \to \Gamma_{\phi_1}(\vec{q}) = 2\pi \sum_{\vec{p}} |M_{\phi_1\to\chi_1\chi_2}|^2 \ \delta(E^{\phi_1}_{\vec{q}} - E^{\chi_1}_{\vec{p}} - E^{\chi_2}_{\vec{q}-\vec{p}}) \,, \tag{6.2.38}$$

is the decay width of ϕ_1.

Therefore in a cascade decay where the intermediate ϕ_1 becomes resonant (near on-shell) and absorbing δE near this resonance into the renormalization of the ϕ_1 energy (mass), the π decay rate (6.2.36) becomes

$$\Gamma_\pi(k) \simeq 2\pi \sum_{\vec{q}} \sum_{\vec{p}} \frac{|M_{\pi\to\phi_1\phi_2}|^2 \, |M_{\phi_1\to\chi_1\chi_2}|^2 \, \delta(E_k^\pi - E_{\vec{k}-\vec{q}}^{\phi_2} - E_{\vec{p}}^{\chi_1} - E_{\vec{q}-\vec{p}}^{\chi_2})}{\left(E_k^\pi - E_{\vec{k}-\vec{q}}^{\phi_2} - E_{\vec{q}}^{\phi_1}\right)^2 + \left(\frac{\Gamma_{\phi_1}(\vec{q})}{2}\right)^2} \tag{6.2.39}$$

This is the usual expression of a decay rate of the parent particle in a cascade in the Breit-Wigner approximation for the propagator of the resonant intermediate state.

Furthermore in the narrow width approximation, (6.2.39) is dominated by the resonance in the denominator and to leading order in the width we can replace

$$\frac{1}{\left(E_k^\pi - E_{\vec{k}-\vec{q}}^{\phi_2} - E_{\vec{q}}^{\phi_1}\right)^2 + \left(\frac{\Gamma_{\phi_1}(\vec{q})}{2}\right)^2} \to \frac{2\pi}{\Gamma_{\phi_1}(\vec{q})} \, \delta\left(E_k^\pi - E_{\vec{k}-\vec{q}}^{\phi_2} - E_{\vec{q}}^{\phi_1}\right) \tag{6.2.40}$$

and write (6.2.39) as

$$\Gamma_\pi(k) \simeq 2\pi \sum_{\vec{q}} |M_{\pi\to\phi_1\phi_2}|^2 \, \delta\left(E_k^\pi - E_{\vec{k}-\vec{q}}^{\phi_2} - E_{\vec{q}}^{\phi_1}\right) \left[\frac{2\pi}{\Gamma_{\phi_1}(\vec{q})} \sum_{\vec{p}} |M_{\phi_1\to\chi_1\chi_2}|^2 \, \delta(E_{\vec{q}}^{\phi_1} - E_{\vec{p}}^{\chi_1} - E_{\vec{q}-\vec{p}}^{\chi_2})\right]. \tag{6.2.41}$$

From (6.2.38) we see that the bracket in this expression equals one, leading to

$$\Gamma_\pi(k) \simeq 2\pi \sum_{\vec{q}} |M_{\pi\to\phi_1\phi_2}|^2 \, \delta\left(E_k^\pi - E_{\vec{k}-\vec{q}}^{\phi_2} - E_{\vec{q}}^{\phi_1}\right) \tag{6.2.42}$$

which is valid in the narrow width approximation. In this case there is no other decay channel for the resonant state and the bracket in (6.2.41) becomes unity, however if there are other decay channels this bracket would be replaced by the branching ratio $BR(\phi_1 \to \chi_1\chi_2)$, which is the usual result for resonant decay in the narrow width approximation.

We are now in position to obtain the time evolution of the amplitudes. In the narrow width limit, the amplitude $\tilde{C}_\pi(\omega)$ given by (6.2.29) features a (narrow) resonance near $\omega \simeq -E_k^\pi$ and is of the Breit-Wigner form

$$\tilde{C}_\pi(\omega \simeq -E_k^\pi) \simeq \frac{-i}{\omega + E_k^\pi - \frac{i}{2}\Gamma_\pi(\vec{k})} \tag{6.2.43}$$

where now E_k^π is the renormalized π single particle energy and Γ_π is the decay width. This is equivalent to a Breit-Wigner approximation to the propagator in terms of a complex pole and as usual ignores the "background" contribution which is perturbative. The time evolution of the amplitude follows straightforwardly, it is given by

$$C_\pi(\vec{k},t) = e^{-iE_k^\pi\, t}\, e^{-\frac{\Gamma_\pi(k)}{2}\, t}\,. \tag{6.2.44}$$

In a similar manner we now obtain the time evolution of $G_{\phi\phi}(t)$, the anti Laplace transform of $\tilde{G}_{\phi\phi}(s)$ in eqn. (6.2.26), with

$$\tilde{G}_{\phi\phi}(s = i\omega + \varepsilon) = \frac{1}{i\omega + iE_{\phi\phi} + i\Sigma^\phi(\vec{k}; i\omega + \varepsilon) + \varepsilon}\,. \tag{6.2.45}$$

It features a complex pole near $\omega \simeq -E_{\phi\phi}$ with

$$i\Sigma^\phi(s = -iE_{\phi\phi} + \varepsilon) = i\Delta E^{\phi_1} + \frac{\Gamma_{\phi_1}}{2} \tag{6.2.46}$$

where the energy shift ΔE^{ϕ_1} and decay width Γ_{ϕ_1} are given by (6.2.37,6.2.38) respectively, and the energy shift ΔE^{ϕ_1} is absorbed into a renormalization of the ϕ_1 mass.

As in the case of $C_\pi(t)$ we now obtain

$$G_{\phi\phi}(t) = e^{-iE_{\phi\phi}\, t}\, e^{-\frac{\Gamma_{\phi_1}}{2}\, t}\,, \tag{6.2.47}$$

where E_{ϕ_1} in $E_{\phi\phi}$ is the renormalized single particle energy for ϕ_1. We now implement the convolution theorem (6.2.25) with $C_\pi(t), G_{\phi\phi}(t)$ given by (6.2.44,6.2.47) respectively and find

$$C_{\phi\phi}(t) = M^*_{\pi\to\phi_1\phi_2}\, \frac{\left[e^{-iE_k^\pi\, t}\, e^{-\frac{\Gamma_\pi(k)}{2}\, t} - e^{-iE_{\phi\phi}\, t}\, e^{-\frac{\Gamma_{\phi_1}}{2}\, t}\right]}{\left(E^\pi - E_{\phi\phi}\right) - \frac{i}{2}\left(\Gamma_\pi - \Gamma_{\phi_1}\right)}\,. \tag{6.2.48}$$

This is the amplitude of the two particle intermediate state with a resonant ϕ_1.

From eqn. (6.2.21,6.2.27) and with the anti Laplace transform of $\tilde{G}_{\phi\chi\chi}$ given by

$$G_{\phi\chi\chi}(t) = e^{-iE_{\phi\chi\chi}\, t} \tag{6.2.49}$$

we find

$$C_{\phi\chi\chi}(t) = -iM^*_{\phi_1\to\chi_1\chi_2} \int_0^t C_{\phi\phi}(t')\, e^{-iE_{\phi\chi\chi}(t-t')}\; dt' \tag{6.2.50}$$

which yields (suppressing all the momenta labels)

$$C_{\phi\chi\chi}(t) = \frac{M^*_{\pi\to\phi_1\phi_2}\, M^*_{\phi_1\to\chi_1\chi_2}\, e^{-iE_{\phi\chi\chi}\,t}}{\Big(E^\pi - E_{\phi\phi}\Big) - \frac{i}{2}\Big(\Gamma_\pi - \Gamma_{\phi_1}\Big)} \times \left\{ \frac{\Big[e^{-i(E^\pi - E_{\phi\chi\chi})t}\, e^{-\frac{\Gamma_\pi}{2}t} - 1\Big]}{(E^\pi - E_{\phi\chi\chi}) - i\frac{\Gamma_\pi}{2}} - \frac{\Big[e^{-i(E_{\phi\phi} - E_{\phi\chi\chi})t}\, e^{-\frac{\Gamma_{\phi_1}}{2}t} - 1\Big]}{(E_{\phi\phi} - E_{\phi\chi\chi}) - i\frac{\Gamma_{\phi_1}}{2}} \right\}. \quad (6.2.51)$$

This is the amplitude of the final three particle state. Although the amplitudes (6.2.48,6.2.51) look unfamiliar, it will be proven in the next section that they satisfy unitarity.

6.3 UNITARITY: POPULATION FLOW.

Unitary time evolution of the state $|\Psi(t)\rangle$ (6.2.4) with the initial condition (6.2.2) implies $\langle\Psi(t)|\Psi(t)\rangle = 1$, namely

$$\sum_n |C_n(t)|^2 = 1 \quad (6.3.1)$$

which for the state (6.2.6) implies

$$|C_\pi(\vec{k},t)|^2 + \sum_{\vec{q}} |C_{\phi\phi}(\vec{k},\vec{q};t)|^2 + \sum_{\vec{q};\vec{p}} |C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t)|^2 + \cdots = 1\,. \quad (6.3.2)$$

The number of π particles with momentum $\vec{k}$ is given by

$$n^\pi_{\vec{k}}(t) = \langle\Psi(t)|a^\dagger_{\pi,\vec{k}}\, a_{\pi,\vec{k}}|\Psi(t)\rangle = |C_\pi(\vec{k};t)|^2 \quad (6.3.3)$$

similarly the number of ϕ_1 particles with particular momentum $\vec{q}$ in the time evolved state $|\Psi(t)\rangle$ is given by

$$n^{\phi_1}_{\vec{q}}(t) = \langle\Psi(t)|a^\dagger_{\phi_1,\vec{q}}\, a_{\phi_1,\vec{q}}|\Psi(t)\rangle = |C_{\phi\phi}(\vec{k},\vec{q};t)|^2 \quad (6.3.4)$$

and the number of *pairs* of $\chi_{1,2}$ particles with momenta $\vec{p}, \vec{q}-\vec{p}$ respectively is

$$n^{\chi_1}_{\vec{p}}(t)\; n^{\chi_2}_{\vec{q}-\vec{p}}(t) = \langle\Psi(t)|\Big(a^\dagger_{\chi_1,\vec{p}}\, a_{\chi_1,\vec{p}}\Big)\Big(a^\dagger_{\chi_2,\vec{q}-\vec{p}}\, a_{\chi_2,\vec{q}-\vec{p}}\Big)|\Psi(t)\rangle = |C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t)|^2 \quad (6.3.5)$$

where $a^\dagger_\alpha$; a_α are the annihilation and creation operators for the quanta of the respective fields. Therefore the probabilities $|C(t)|^2$ are also a measure of the *population* of the many particle states upon decay of the initial state.

Since $|C_\pi(\vec{k},0)| = 1$; $|C_{\phi\phi}(\vec{k},\vec{q};0)| = 0$; $|C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};0)| = 0$, as time evolves the probabilities $|C_{\phi\phi}(\vec{k},\vec{q};t)|^2$ of the intermediate (resonant) state $\left|\phi_{1,\vec{q}}\phi_{2,\vec{k}-\vec{q}}\right\rangle$ will grow as the $|\pi\rangle$ state decays therefore producing ϕ_1 particles, however these particles eventually decay into final state particles $\left|\phi_{2,\vec{k}-\vec{q}}\chi_{1,\vec{p}}\chi_{2,\vec{q}-\vec{p}}\right\rangle$ whose population $|C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t)|^2$ will grow in time . Therefore we expect the physical picture: the amplitude of the initial state decays, while the amplitude of the intermediate resonant state grows at early times eventually decays into $\chi_{1,2}$ particles, and the amplitude of the final state grows more slowly than the intermediate state at early times, but it reaches an asymptotic final value to fulfill the unitarity relation (6.3.2) with $|C_\pi(\vec{k},\infty)| = 0$; $|C_{\phi\phi}(\vec{k},\vec{q};\infty)| = 0$; $|C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};\infty)| \neq 0$ with

$$\sum_{\vec{q}}\sum_{\vec{p}} |C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};\infty)|^2 = 1\,. \tag{6.3.6}$$

This physical picture describes probability or *population flow*, namely that the probabilities of the various states evolve in time in such a way that the $\langle\Psi(t)|\Psi(t)\rangle$ is constant but population and probability flows among multiparticle states and at asymptotically long times only the stable final particle states feature non-vanishing amplitudes.

The main goal of this section is to understand how unitarity is manifest in the time evolution of the probabilities. At first this notion seems puzzling: at $t = 0$ the initial state has unit probability and the intermediate and final states vanishing probability. The matrix elements connecting the initial, intermediate and final states are all perturbatively small in the couplings, yet at asymptotically late time when the initial state has decayed, the total probability of the final state, related to the initial state by perturbative matrix elements must be unity, highlighting the non-perturbative aspects of the dynamics.

We now study this process in detail to analyze the various time scales associated with this "population flow". Let us introduce

$$\mathcal{E} = E_{\phi\phi} - E^\pi_k \quad ; \quad \Delta\Gamma = \Gamma_\pi - \Gamma_{\phi_1}\,, \tag{6.3.7}$$

in terms of which

$$|C_{\phi\phi}(\vec{k},\vec{q};t)|^2 = |M_{\pi\to\phi_1\phi_2}|^2\, e^{-\Gamma_{\phi_1}\, t}\, \frac{\left|e^{i\mathcal{E}t}\; e^{-\frac{\Delta\Gamma}{2}\,t} - 1\right|^2}{\left[\mathcal{E}^2 + \left(\frac{\Delta\Gamma}{2}\right)^2\right]}\,. \tag{6.3.8}$$

In the narrow width limit this expression becomes proportional to $\delta(\mathcal{E})/|\Delta\Gamma|$, in order to find the proportionality factor we integrate (6.3.8) in $-\infty \leq \mathcal{E} \leq \infty$ and obtain to leading order in the narrow width limit

$$|C_{\phi\phi}(\vec{k},\vec{q};t)|^2 = 2\pi\; |M_{\pi\to\phi_1\phi_2}|^2\, \frac{\left(e^{-\Gamma_{\phi_1}\, t} - e^{-\Gamma_\pi\, t}\right)}{\Gamma_\pi - \Gamma_{\phi_1}}\; \delta(E^\pi_k - E^{\phi_1}_{\vec{q}} - E^{\phi_2}_{\vec{k}-\vec{q}})\,. \tag{6.3.9}$$

This result is obviously non-perturbative: while in the numerator is $|M_{\pi\to\phi_1\phi_2}|^2 \propto g^2_\pi$ the denominator is *also* $\propto g^2_\pi, g^2_\phi$ exhibiting the non-perturbative nature of the result. At very early times $t \ll 1/\Gamma_\pi; 1/\Gamma_{\phi_1}$ the total number of ϕ_1 particles grows linearly with time with a rate

$$\Gamma(\pi \to \phi_1\phi_2) = 2\pi \sum_{\vec{q}}\; |M_{\pi\to\phi_1\phi_2}|^2 \delta(E^\pi_k - E^{\phi_1}_{\vec{q}} - E^{\phi_2}_{\vec{k}-\vec{q}}) \tag{6.3.10}$$

however it reaches a maximum at $t = t^*$ given by

$$t^* = \frac{\ln\left[\frac{\Gamma_\pi}{\Gamma_{\phi_1}}\right]}{\Gamma_\pi - \Gamma_{\phi_1}} \tag{6.3.11}$$

and falls off exponentially on a time scale determined by the smaller of $\Gamma_\pi, \Gamma_{\phi_1}$.

In particular if the π particle decays at rest the total number of ϕ_1 particles is given by

$$N_{\phi_1}(t) = \sum_{\vec{q}} |C_{\phi\phi}(\vec{k},\vec{q};t)|^2 = \frac{\Gamma(\pi\to\phi_1\phi_2)}{\Gamma_\pi}\left[\frac{e^{-\Gamma^*_{\phi_1}\, t} - e^{-\Gamma_\pi\, t}}{1 - \frac{\Gamma^*_{\phi_1}}{\Gamma_\pi}}\right] \tag{6.3.12}$$

where

$$\Gamma^*_{\phi_1} = \Gamma_{\phi_1}(q*)\;\;;\;\; q* = \frac{1}{2m_\pi}\Big[m^4_\pi + m^4_{\phi_1} + m^4_{\phi_2} - 2m^2_\pi\, m^2_{\phi_1} - 2m^2_\pi\, m^2_{\phi_2} - 2m^2_{\phi_1}\, m^2_{\phi_2}\Big]^{\frac{1}{2}}\,. \tag{6.3.13}$$

The time dependent function in the bracket in (6.3.12) that determines the population of the resonant particles is shown in fig. (31) as a function of $\Gamma_\pi\, t$ for $\Gamma_{\phi_1}/\Gamma_\pi = 0.1, 10$, displaying the behavior discussed above: an early linear growth from the production of resonant states

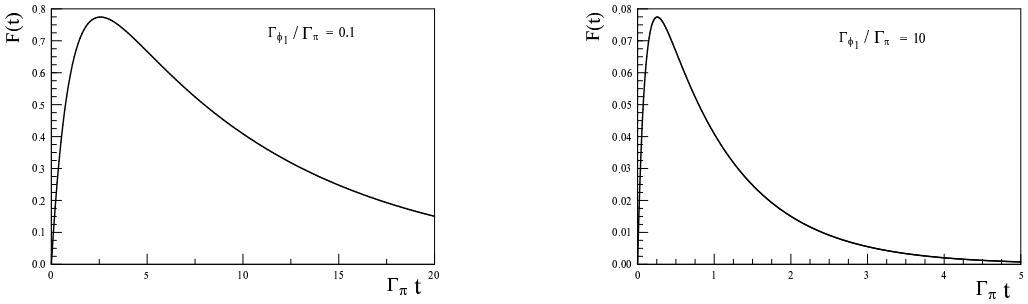


Figure 31: The function $F(t) = \left(e^{-\Gamma_{\phi_1} t} - e^{-\Gamma_\pi t}\right)/(1 - \Gamma_{\phi_1}/\Gamma_\pi)$ for $\Gamma_{\phi_1}/\Gamma_\pi = 0.1, 10$.

from the decay of the parent particle for $t \ll \Gamma_\pi, \Gamma_{\phi_1}$ growing to a maximum and falling off exponentially on the longer time scale determined by the smaller of the decay widths.

We now turn to $|C_{\phi\chi\chi}(\vec{k}, \vec{q}, \vec{p}; t)|^2$ focusing first on the asymptotic $t \to \infty$ limit, we find

$$\sum_{\vec{q},\vec{p}} |C_{\phi\chi\chi}(\vec{k}, \vec{q}, \vec{p}; \infty)|^2 = \sum_{\vec{q},\vec{p}} \frac{|M_{\pi\to\phi_1\phi_2}|^2 \ |M_{\phi_1\to\chi_1\chi_2}|^2}{\left[\left(E_{\phi\chi\chi} - E_k^\pi\right)^2 + \left(\frac{\Gamma_\pi}{2}\right)^2\right]\left[\left(E_{\phi\chi\chi} - E_{\phi\phi}\right)^2 + \left(\frac{\Gamma_{\phi_1}}{2}\right)^2\right]} \,. \tag{6.3.14}$$

Remarkably this expression is very similar to the asymptotic limit of the probability of a two photon state in a two-level radiative cascade in quantum optics[204, 232].

In the narrow width limit this expression is dominated by the resonances and as usual we approximate

$$\frac{1}{\left[\left(E_{\phi\chi\chi} - E_k^\pi\right)^2 + \left(\frac{\Gamma_\pi}{2}\right)^2\right]} \to \frac{2\pi}{\Gamma_\pi}\,\delta\left(E_{\phi\chi\chi} - E_k^\pi\right) \tag{6.3.15}$$

therefore

$$\sum_{\vec{q},\vec{p}} |C_{\phi\chi\chi}(\vec{k}, \vec{q}, \vec{p}; \infty)|^2 = \frac{2\pi}{\Gamma_\pi} \sum_{\vec{q},\vec{p}} \frac{|M_{\pi\to\phi_1\phi_2}|^2 \ |M_{\phi_1\to\chi_1\chi_2}|^2 \ \delta\left(E_{\phi\chi\chi} - E_k^\pi\right)}{\left[\left(E_k^\pi - E_{\phi\phi}\right)^2 + \left(\frac{\Gamma_{\phi_1}}{2}\right)^2\right]} \tag{6.3.16}$$

where in the denominator we used the delta function to set $E_{\phi\chi\chi} = E_k^\pi$.

Upon comparing this result with the result for Γ_π given by (6.2.39) (see the definitions (6.2.10,6.2.11) we find

$$\sum_{\vec{q},\vec{p}} |C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};\infty)|^2 = 1 \tag{6.3.17}$$

which is the manifestation of unitarity in the asymptotic limit.

The analysis for finite time is lengthy and relegated to appendix (6.7.1), here we present the main results.

In the narrow width limit we replace

$$\frac{1}{\left[\left(E_{\phi\phi}-E_k^\pi\right)^2+\left(\frac{\Gamma_{\phi_1}}{2}\right)^2\right]} \to \frac{2\pi}{\Gamma_{\phi_1}}\,\delta\Big(E_{\phi\phi}-E_k^\pi\Big) \tag{6.3.18}$$

and to leading order in the widths it follows that

$$\sum_{\vec{q},\vec{p}} |C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};\infty)|^2 = (2\pi)^2 \sum_{\vec{q},\vec{p}} \frac{|M_{\pi\to\phi_1\phi_2}|^2\ |M_{\phi_1\to\chi_1\chi_2}|^2}{\Gamma_\pi\ \Gamma_{\phi_1}}\ \delta\Big(E_{\phi\chi\chi}-E_k^\pi\Big)\ \delta\Big(E_{\phi\phi}-E_k^\pi\Big)\,, \tag{6.3.19}$$

combining this expression with (6.7.3) and (6.7.11) we finally find

$$\begin{aligned}\sum_{\vec{q},\vec{p}} |C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t)|^2 &= (2\pi)^2 \sum_{\vec{q},\vec{p}} \frac{|M_{\pi\to\phi_1\phi_2}|^2\ |M_{\phi_1\to\chi_1\chi_2}|^2}{\Gamma_\pi\ \Gamma_{\phi_1}\ (\Gamma_\pi-\Gamma_{\phi_1})} \\ \Big\{\Gamma_\pi\ (1-e^{-\Gamma_{\phi_1}t})-\Gamma_{\phi_1}\ (1-e^{-\Gamma_\pi t})\Big\} &\times \delta\Big(E_{\phi\chi\chi}-E_k^\pi\Big)\ \delta\Big(E_{\phi\phi}-E_k^\pi\Big)\,.\end{aligned} \tag{6.3.20}$$

This expression may be simplified by realizing that $\Gamma_{\phi_1}(\vec{q})$ depends on $\vec{q}$ but not on $\vec{p}$ and because of the second delta function in (6.3.20) we can set $\delta\Big(E_{\phi\chi\chi}-E_k^\pi\Big) \to \delta\Big(E_{\phi\chi\chi}-E_{\phi\phi}\Big) = \delta(E_{\vec{q}}^{\phi_1}-E_{\vec{p}}^{\chi_1}-E_{\vec{q}-\vec{p}}^{\chi_2})$ in the first delta function, using the result (6.2.38) we finally find

$$\begin{aligned}\sum_{\vec{q},\vec{p}} |C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t)|^2 &= \frac{2\pi}{\Gamma_\pi}\sum_{\vec{q}} |M_{\pi\to\phi_1\phi_2}|^2 \\ &\times \left[\frac{\Gamma_\pi\ (1-e^{-\Gamma_{\phi_1}t})-\Gamma_{\phi_1}\ (1-e^{-\Gamma_\pi t})}{(\Gamma_\pi-\Gamma_{\phi_1})}\right]\delta\Big(E_k^\pi-E_{\vec{k}-\vec{q}}^{\phi_2}-E_{\vec{q}}^{\phi_1}\Big)\,.\end{aligned} \tag{6.3.21}$$

Combining this result with (6.3.9) and using (6.2.41) we find

$$\sum_{\vec{q}} |C_{\phi\phi}(\vec{k},\vec{q};t)|^2 + \sum_{\vec{q},\vec{p}} |C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t)|^2 = 1 - e^{-\Gamma_\pi t} = 1 - |C_\pi(t)|^2 \,, \tag{6.3.22}$$

this is the final result confirming unitary time evolution and the flow of population from the initial through the intermediate to the final state.

Therefore an important conclusion that follows from this analysis is that the amplitudes (6.2.48,6.2.50) are the correct ones insofar as they manifestly satisfy unitarity.

In particular when the π particle decays at rest, from the result (6.3.22) and the identification (6.3.5) the total number of *pairs* of χ_1, χ_2 particles is given by

$$N_{\chi\chi}(t) = \sum_{\vec{q},\vec{p}} |C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t)|^2 = \left[\frac{\Gamma_\pi \, (1 - e^{-\Gamma^*_{\phi_1} t}) - \Gamma^*_{\phi_1} \, (1 - e^{-\Gamma_\pi t})}{(\Gamma_\pi - \Gamma^*_{\phi_1})} \right] \tag{6.3.23}$$

where $\Gamma^*_{\phi_1}$ is given by (6.3.13). In summary:

$$N_\pi(t) = e^{-\Gamma_\pi t} \tag{6.3.24}$$

$$N_{\phi_1}(t) = \frac{e^{-\Gamma^*_{\phi_1} t} - e^{-\Gamma_\pi t}}{1 - \frac{\Gamma^*_{\phi_1}}{\Gamma_\pi}} \tag{6.3.25}$$

$$N_{\chi\chi}(t) = \frac{\Gamma_\pi \, (1 - e^{-\Gamma^*_{\phi_1} t}) - \Gamma^*_{\phi_1} \, (1 - e^{-\Gamma_\pi t})}{(\Gamma_\pi - \Gamma^*_{\phi_1})} \,, \tag{6.3.26}$$

with

$$N_\pi(t) + N_{\phi_1}(t) + N_{\chi\chi}(t) = 1 \,. \tag{6.3.27}$$

The populations are shown in fig. (32) as a function of time for $\Gamma_{\phi_1}/\Gamma_\pi = 0.1, 10$.

Discussion: The expressions (6.3.24-6.3.26) and the figures (31,32) display the behavior of the populations of resonant and final states and describe the main time dependent physical phenomena of the cascade decay. As discussed above at early time $t \ll 1/\Gamma_\pi, 1/\Gamma_{\phi_1}$ the number of resonances grows linearly in t as $N_{\phi_1}(t) \simeq \Gamma_\pi t$, as the decay the the parent particle increases the population of resonances, riches a maximum at t^* given by (6.3.11) and decays on the longer time scale. However the population of the final state $\chi\chi\phi_2$ at early time grows as $N_{\chi\chi}(t) \simeq \Gamma_\pi \, \Gamma_\phi t^2/2$, namely much *slower* than the population of the intermediate resonant state. This is a result of a second order process as the build-up of the

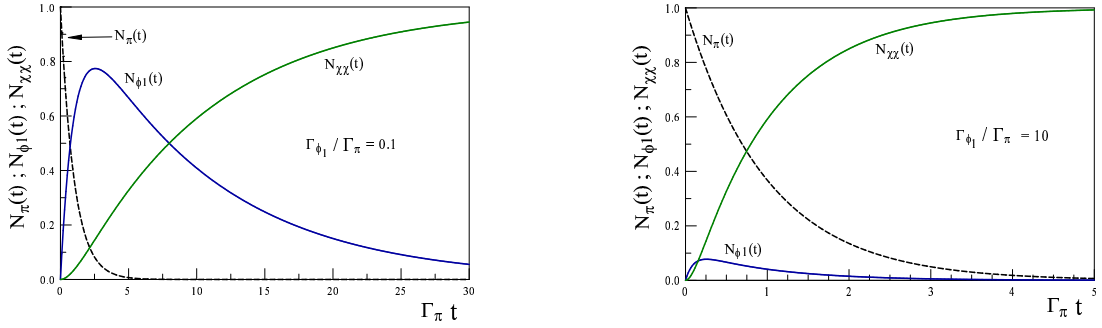


Figure 32: The number of π-particles $N_\pi(t)$, resonant states $N_{\phi_1}(t)$ and χ-pairs $N_{\chi\chi}(t)$ vs. $\Gamma_\pi\, t$ for $\Gamma_{\phi_1}/\Gamma_\pi = 0.1, 10$.

final state requires first to build up the population of the resonant state from the decay of the parent state as $\simeq \Gamma_\pi\, t$ and the build-up of the final state from the decay of the *populated* intermediate resonant state as $\Gamma_{\phi_1}\, t$. The behavior of the populations at intermediate and long times depends on the particular cases $\Gamma_\pi \gg \Gamma_{\phi_1}$ and $\Gamma_\pi \ll \Gamma_{\phi_1}$.

- $\mathbf{\Gamma_\pi \gg \Gamma_{\phi_1}}$: In this case the slowly decaying intermediate resonant state acts as a "bottleneck", the rapid decay of the parent particle builds up the population of the resonant states which grows to a maximum at t^* with an amplitude $N_{\phi_1}(t^*)$. For $\Gamma_\pi \gg \Gamma_{\phi_1}$ it follows that $e^{-\Gamma_\pi\, t^*} \ll 1$; $e^{-\Gamma_{\phi_1}\, t^*} \sim 1$ and $N_{\phi_1}(t^*) \sim 1$. The decay of the resonant state into the final state particles occurs on a *much slower* time scale of the order of $1/\Gamma_{\phi_1}$. In this limit the cascade decay can be described *sequentially* as $\pi \to \phi_2\phi_1$; $\phi_2\phi_1 \to \chi_1\chi_2\phi_2$ where the intermediate resonant state attains an amplitude $\simeq 1$, nearly saturating unitarity on a short time scale $t \simeq t^* = \ln[\Gamma_\pi/\Gamma_{\phi_1}]/(\Gamma_\pi - \Gamma_{\phi_1})$ and decays slowly on the time scale $1/\Gamma_{\phi_1}$. For $t > 1/\Gamma_\pi$ it follows that the population of the final state $N_{\chi\chi}(t) \simeq (1 - e^{-\Gamma^*_{\phi_1}\, t})$ consistently with the decay of an initial resonant state of nearly unit amplitude.
- $\mathbf{\Gamma_\pi \ll \Gamma_{\phi_1}}$: In this case the intermediate resonant state decays on time scales shorter

than that of the parent particle, as a result there is very little population build-up of the resonant state and the initial population of the parent particle is "transferred" directly to the final state on a time scale $\simeq 1/\Gamma_\pi$. In this limit $\Gamma_{\phi_1}\, t^* \gg 1$ and $N_{\phi_1}(t^*) \simeq e^{-\Gamma_\pi\, t^*}\ \Gamma_\pi/\Gamma_{\phi_1} \ll 1$, and for $t \gtrsim t^*$ it follows that $N_{\chi\chi}(t) \simeq (1 - e^{-\Gamma_\pi\, t})$ describing the build up of the population of the final states directly from the decay of the parent particle. In this limit the decay of the parent particle can be described as a direct decay into the final states as the intermediate resonant state is so short-lived that the population of resonances does not build up substantially.

- $\boldsymbol{\Gamma_\pi = \Gamma_{\phi_1}}$: in this (unlikely) case it is straightforward to find

$$N_{\phi_1}(t) = \Gamma_\pi\, t\ e^{-\Gamma_\pi\, t}\ \ ;\ \ N_{\chi\chi}(t) = 1 - e^{-\Gamma_\pi\, t}\Big[1 + \Gamma_\pi\, t\Big]\ \ ;\ \ t^* = 1/\Gamma_\pi\,. \tag{6.3.28}$$

The "bottleneck" in the case when $\boldsymbol{\Gamma_\pi \gg \Gamma_{\phi_1}}$ is reminiscent of a similar phenomenon in the production of Helium during Big Bang Nucleosynthesis (BBN) via the formation of a deuteron bound state in $n-p$ collisions. The deuteron is photo-dissociated by the high energy photons in the blackbody tail and does not form until the ambient temperature falls to about 80 keV resulting in a delayed transition from the initial $n-p$ to the final state. However, although there is a similarity in that the decay to the final state is delayed, the physical reasons are different: in the case under consideration in this article, the delayed decay to the final state is a consequence of the on-shell formation of a long-lived resonant state, whereas in BBN the delay is a consequence of photodissociation of the deuteron intermediate state which does not form until the temperature falls well below the binding energy of the bound state ≈ 2 MeV. Once the deuteron is formed the fusion reactions end up in ^4He very fast.

Generality of the results:

Although we focused on the specific example given by the interaction Hamiltonian (6.2.1) the procedure leading to the equations for the amplitudes is general. In particular we have left the matrix elements (6.2.8,6.2.9) indicated without using their explicit form and the final results for the amplitudes depend solely on the transition matrix elements, decay widths of the parent and resonant states and single particle energies. Therefore the extension of the

above results to any other theory in which a cascade decay proceeds via an intermediate resonant state is straightforward and can be described by the above results by replacing the proper matrix elements and decay widths.

6.4 FIELD THEORETIC GENERALIZATION OF WIGNER-WEISSKOPF:

The Wigner-Weisskopf theory of spontaneous emission[100] plays an important role in quantum optics[204, 224, 212] and in particle physics it is one of the main approaches to study the dynamics of the $K^0 - \overline{K^0}$ system[230, 101] (also $B^0 - \overline{B^0}$). A quantum field theoretical generalization of this important method was provided in ref.[231] with an extension in cosmology in refs.[175, 311]. However, to the best of our knowledge the method has not been extended to the case of radiative cascades. This is the goal of this section.

Again, writing the Hamiltonian as $H = H_0 + H_I$, and passing to the interaction picture instead of using the Schrodinger picture, as in (6.2.3), the time evolution of the quantum state is given by

$$i\frac{d}{dt}|\Psi(t)\rangle_I = H_I(t)|\Psi(t)\rangle_I \tag{6.4.1}$$

where $H_I(t)$ is the interaction Hamiltonian in the interaction picture.

The state $|\Psi(t)\rangle_I$ is then expanded in the basis of free particle Fock states $|m\rangle$ eigenstates of H_0, namely

$$|\Psi(t)\rangle_I = \sum_m A_m(t)|m\rangle\,. \tag{6.4.2}$$

Using the orthogonality relation on any state expanded in this fashion leads to the following relation

$$\dot{A}_n(t) = -i\sum_m \langle n|H_I(t)|m\rangle A_m(t)\,. \tag{6.4.3}$$

This allows for a *time dependent* interaction Hamiltonian, a situation common to field theories in curved spacetime[175, 311]. Specifically, the method described below, when gen-

eralized to the case of an expanding cosmology is amenable to be implemented to study the cascade decay of inflationary quanta discussed in ref.[311]. We intend to apply the method developed here to the case of the cosmological cascade decay in future studies.

For the purposes of this work, ensuring that the Wigner-Weisskopf procedure reproduces the results obtained via Laplace transform will confirm that this real time method provides a non-perturbative resummation that yields the correct time evolution at least in the cases where it can be compared to known results.

In the interaction picture the quantum state that describes the cascade decay is given by

$$|\Psi(t)\rangle_I = A_\pi(\vec{k},t)\big|\pi_{\vec{k}}\rangle + \sum_{\vec{q}} A_{\phi\phi}(\vec{k},\vec{q};t)\,\big|\phi_{1,\vec{q}}\phi_{2,\vec{k}-\vec{q}}\rangle + \sum_{\vec{q};\vec{p}} A_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t)\,\big|\phi_{2,\vec{k}-\vec{q}}\chi_{1,\vec{p}}\chi_{2,\vec{q}-\vec{p}}\rangle + \cdots \tag{6.4.4}$$

When the interaction Hamiltonian in the Schroedinger picture is time independent as is the case in Minkowski space time in absence of external sources, the coefficients in this expression differ from those in Section 6.2 by the relation $C = Ae^{-iEt}$, where E is the energy (eigenvalue of H_0) of the particular state in the expansion, this is the difference between the two pictures.

Restricting attention to Minkowski spacetime in absence of explicit time dependent sources, the orthogonality relations lead to matrix elements of the interaction Hamiltonian given by

$$\mathcal{M}_{mn}(t) = \langle m|H_I(t)|n\rangle = e^{i(E_m-E_n)t}\langle m|H_I|n\rangle \quad ; \quad H_I(t) = e^{iH_0t}\,H_I\,e^{-iH_0t}. \tag{6.4.5}$$

Again, consider the interaction Hamiltonian given by (6.2.1) and the initial conditions given by (6.2.2). To simplify notation, the definitions introduced in Eqs (6.2.8-6.2.11) will be used leading to the following equations for the coefficients

$$\dot{A}_\pi(\vec{k},t) = -i\sum_{\vec{q}} M_{\pi\to\phi_1\phi_2}A_{\phi\phi}(\vec{k},\vec{q},t)\,e^{i(E^\pi_k-E_{\phi\phi})t} \quad ; \quad A_\pi(\vec{k},0) = 1\,, \tag{6.4.6}$$

$$\dot{A}_{\phi\phi}(\vec{k},\vec{q},t) = -\,iM^{*}_{\pi\to\phi_1\phi_2}A_{\pi}(\vec{k},t)\,e^{i(E_{\phi\phi}-E^{\pi}_{k})t} \\ -\,i\sum_{\vec{p}} M_{\phi_1\to\chi_1\chi_2}A_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p},t)\,e^{i(E_{\phi\phi}-E_{\phi\chi\chi})t} \quad ; \quad A_{\phi\phi}(\vec{k},\vec{p},0)=0\,, \tag{6.4.7}$$

$$\dot{A}_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p},t) = -iM^{*}_{\phi_1\to\chi_1\chi_2}e^{-i(E_{\phi\phi}-E_{\phi\chi\chi})t}A_{\phi\phi}(\vec{k},\vec{q},t) \quad ; \quad A_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p},0)=0\,. \tag{6.4.8}$$

Again we solve the hierarchy from the bottom up. The solution of (6.4.8) is

$$A_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p},t) = -iM^{*}_{\phi_1\to\chi_1\chi_2}\int_0^t e^{-i(E_{\phi\phi}-E_{\phi\chi\chi})t'}A_{\phi\phi}(\vec{k},\vec{q},t')dt' \tag{6.4.9}$$

and combining (6.4.7) and (6.4.9) yields (now suppressing the momenta in the arguments)

$$\dot{A}_{\phi\phi}(t)+\sum_{\vec{p}}|M_{\phi_1\to\chi_1\chi_2}|^2\int_0^t dt' e^{-i(E_{\phi\phi}-E_{\phi\chi\chi})(t'-t)}A_{\phi\phi}(t') = -iM^{*}_{\pi\to\phi_1\phi_2}e^{-i(E_{\pi}-E_{\phi\phi})t}A_{\pi}(t)\,. \tag{6.4.10}$$

A perturbative solution to this integro-differential equation in terms of A_π is straightforward, however, it leads to resonant denominators and its eventual breakdown. Instead we implement a non-perturbative approach that provides a resummation that incorporates consistently the width of the resonant state.

In order to implement this method to solve (6.4.10), let us first focus on the homogeneous case neglecting the right hand side. Consider

$$\dot{A}^{H}_{\phi\phi}(t)+\sum_{\vec{p}}|M_{\phi_1\to\chi_1\chi_2}|^2\int_0^t dt' e^{-i(E_{\phi\phi}-E_{\phi\chi\chi})(t'-t)}A^{H}_{\phi\phi}(t') = 0 \quad ; \quad A^{H}_{\phi\phi}(0)=1\,. \tag{6.4.11}$$

This equation simplifies by implementing a Markovian approximation which is justified in weak coupling. Since the term inside the integrand $|M_{\phi_1\to\chi_1\chi_2}|^2\sim g^2_\phi \ll 1$ it follows that

$\dot{A}^H_{\phi\phi} \propto g^2_\phi \ll 1$, namely the amplitudes vary *slowlya* in time. To clearly see the nature of the approximation, the following is introduced

$$W^\phi_0(t,t') = \sum_{\vec{p}} |M_{\phi_1 \to \chi_1\chi_2}|^2 \int_0^{t'} dt'' e^{-i(E_{\phi\phi} - E_{\phi\chi\chi})(t''-t)} \tag{6.4.12}$$

which has the properties

$$\frac{d}{dt'} W^\phi_0(t,t') = \sum_{\vec{p}} |M_{\phi_1 \to \chi_1\chi_2}|^2 e^{-i(E_{\phi\phi} - E_{\phi\chi\chi})(t'-t)} \sim \mathcal{O}(g^2_\phi) \;\; ; \;\; W^\phi_0(t,0) = 0 \tag{6.4.13}$$

An integration by parts produces

$$\int_0^t dt' \frac{d}{dt'} W^\phi_0(t,t') A_{\phi\phi}(t') = W^\phi_0(t,t) A_{\phi\phi}(t) - \int_0^t dt' \dot{A}_{\phi\phi}(t') W^\phi_0(t,t') \tag{6.4.14}$$

This can be repeated systematically, producing higher order derivatives by using the natural definition

$$W^\phi_N(t,t') = \int_0^{t'} dt'' W^\phi_{N-1}(t,t'') \;\; ; \;\; W^\phi_N(t,0) = 0\,. \tag{6.4.15}$$

Repeated integration by parts produces the series

$$\int_0^t dt' \frac{d}{dt'} W^\phi_0(t,t') A_{\phi\phi}(t') = W^\phi_0(t,t) A_{\phi\phi}(t) - W^\phi_1(t,t) \dot{A}_{\phi\phi}(t) + W^\phi_2(t,t) \ddot{A}_{\phi\phi}(t) + ... \tag{6.4.16}$$

where each term has a multiplicative factor $W^\phi_N \sim g^2_\phi$ and $\dot{A} \propto g^2_\phi; \ddot{A} \propto g^4_\phi$ etc. Truncating this series to leading order, namely keeping only W^ϕ_0 gives

$$\dot{A}^H_{\phi\phi} + W^\phi_0(t,t) A^H_{\phi\phi}(t) = 0 \;\; ; \;\; A^H_{\phi\phi}(0) = 1\,. \tag{6.4.17}$$

This makes it apparent that the lowest order solution results from keeping only W_0, this is the Markovian approximation. So to lowest order, the homogenous solution is written

$$A^H_{\phi\phi}(t) = e^{-\int_0^t dt' W^\phi_0(t',t')} \tag{6.4.18}$$

where

$$\int_0^t dt' W_0^{\phi}(t',t') = it \sum_{\vec{p}} \frac{|M_{\phi_1\to\chi_1\chi_2}|^2}{(E_{\phi\phi}-E_{\phi\chi\chi})} \left\{ \left(1 - \frac{\sin(E_{\phi\phi}-E_{\phi\chi\chi})\,t}{(E_{\phi\phi}-E_{\phi\chi\chi})\,t}\right) -i \left(\frac{1-\cos(E_{\phi\phi}-E_{\phi\chi\chi})\,t}{(E_{\phi\phi}-E_{\phi\chi\chi})t}\right)\right\} \tag{6.4.19}$$

We prove in appendix (6.7.2) (see also ref.[175]) that in the long time limit for time scales that are much larger than the energy uncertainty ($t \gg 1/(E_{\phi\phi}-E_{\phi\chi\chi})$),

$$\int_0^t dt' W_0^{\phi}(t',t') \to t \left(i\mathcal{P} \sum_{\vec{p}} \frac{|M_{\phi_1\to\chi_1\chi_2}|^2}{E_{\phi\phi}-E_{\phi\chi\chi}} + \pi \sum_{\vec{q}} |M_{\phi_1\to\chi_1\chi_2}(p,q)|^2 \delta(E_{\phi\phi}-E_{\phi\chi\chi}) \right) \tag{6.4.20}$$

However, the same result can be obtained from replacing $E_{\phi\phi}-E_{\phi\chi\chi} \to E_{\phi\phi}-E_{\phi\chi\chi}+i\varepsilon$ with $\varepsilon \to 0^+$ and $t=t' \to \infty$ in (6.4.12) which yields

$$\lim_{t\to\infty} \int_0^t dt' W_0(t',t') = i\,t \sum_{\vec{p}} \frac{|M_{\phi_1\to\chi_1\chi_2}|^2}{(E_{\phi\phi}-E_{\phi\chi\chi}+i\varepsilon)} = \left(i\Delta E^{\phi_1} + \frac{\Gamma_{\phi_1}}{2}\right) t \tag{6.4.21}$$

where ΔE^{ϕ_1} and Γ_{ϕ_1} are given by (6.2.37,6.2.38) respectively, and the homogenous solution becomes

$$A^H_{\phi\phi}(t) = e^{-i\,\Delta E^{\phi_1} t}\; e^{-\frac{\Gamma_{\phi_1}}{2} t}\,. \tag{6.4.22}$$

Now with the left hand side of eqn. (6.4.10) replaced by the Markovian approximation (6.4.17) and using the above result, the full solution of (6.4.10) is given by

$$A_{\phi\phi}(t) = -iM^*_{\pi\to\phi_1\phi_2} e^{-i(\Delta E_{\phi_1} - i\frac{\Gamma_{\phi_1}}{2})\,t} \int_0^t dt'\, A_\pi(t')\, e^{-i(E^\pi - E_{\phi\phi} - \Delta E_{\phi_1} + i\frac{\Gamma_{\phi_1}}{2})\,t'}\,. \tag{6.4.23}$$

Inserting this solution into Eq (6.4.6) we obtain

$$\dot{A}_\pi(t) + \sum_{\vec{q}} |M_{\pi\to\phi_1\phi_2}|^2 \int_0^t dt' e^{i(E^\pi - E_{\phi\phi} - \Delta E_{\phi_1} + i\frac{\Gamma_{\phi_1}}{2})(t-t')} A_\pi(t') = 0 \;\;;\;\; A_\pi(0) = 1\,. \tag{6.4.24}$$

At this stage, we implement again a Markovian approximation as described above, which is justified in this case by the weak coupling $g_\pi \ll 1$ so that the integrand can be treated as slowly varying and carrying out the same expansion as above and keeping the lowest order we obtain

$$\dot{A}_\pi(t) + W_0^\pi(t,t)A_\pi(t) = 0 \;\; ; \;\; A_\pi(0) = 1\,, \tag{6.4.25}$$

where

$$W_0^\pi(t,t') = \sum_{\vec{q}} |M_{\pi\to\phi_1\phi_2}|^2 \int_0^{t'} dt''\, e^{-i(E^\pi - E_{\phi\phi} - \Delta E_{\phi_1} + i\frac{\Gamma_{\phi_1}}{2})(t''-t)} \tag{6.4.26}$$

and in the same manner as before the solution is now given by

$$A_\pi(t) = e^{-\int_0^t dt' W_0^\pi(t',t')} \tag{6.4.27}$$

where again taking the long time limit

$$\int_0^t dt' W_0^\pi(t',t') = i\,t \sum_{\vec{q}} \frac{|M_{\pi\to\phi_1\phi_2}|^2}{E^\pi - E_{\phi\phi} - \Delta E_{\phi_1} + i\frac{\Gamma_{\phi_1}}{2} + i\varepsilon}\,, \tag{6.4.28}$$

where for $\Gamma_{\phi_1} \neq 0$ one can neglect ε. The solution now becomes

$$A_\pi(t) = e^{-i\Delta E^\pi\, t} e^{-\frac{\Gamma_\pi}{2}\, t} \tag{6.4.29}$$

where

$$\Delta E^\pi = \sum_{\vec{q}} \frac{|M_{\pi\to\phi_1\phi_2}|^2 \left(E^\pi - E_{\phi\phi} - \Delta E_{\phi_1}\right)}{\left(E^\pi - E_{\phi\phi} - \Delta E_{\phi_1}\right)^2 + \left(\frac{\Gamma_{\phi_1}}{2}\right)^2} \tag{6.4.30}$$

$$\Gamma_\pi(k) = \sum_{\vec{q}} \frac{|M_{\pi\to\phi_1\phi_2}|^2\, \Gamma_{\phi_1}}{\left(E^\pi - E_{\phi\phi} - \Delta E_{\phi_1}\right)^2 + \left(\frac{\Gamma_{\phi_1}}{2}\right)^2}\,. \tag{6.4.31}$$

This is now used to solve Eq (6.4.7) resulting in the following expression

$$A_{\phi\phi}(t) = -iM^*_{\pi\to\phi_1\phi_2} e^{-i(\Delta E_{\phi_1} - i\frac{\Gamma_{\phi_1}}{2})\,t} \int_0^t dt' e^{-i(E^\pi - E_{\phi\phi} + \Delta E^\pi - \Delta E_{\phi_1} - \frac{i}{2}(\Gamma_\pi - \Gamma_{\phi_1}))\,t'}$$
$$= \frac{M^*_{\pi\to\phi_1\phi_2} e^{-i(\Delta E_{\phi_1} - i\frac{\Gamma_{\phi_1}}{2})\,t}}{E^\pi - E_{\phi\phi} + \Delta E^\pi - \Delta E_{\phi_1} - \frac{i}{2}(\Gamma_\pi - \Gamma_{\phi_1})} \left[e^{-i(E^\pi - E_{\phi\phi} + \Delta E^\pi - \Delta E_{\phi_1} - \frac{i}{2}(\Gamma_\pi - \Gamma_{\phi_1}))t} - 1 \right] \tag{6.4.32}$$

Absorbing the energy shifts into renormalizations of the mass as before, namely

$$E^\pi = E^\pi_0 + \Delta E^\pi \quad ; \quad E_{\phi\phi} = E^0_{\phi_1} + \Delta E_\phi + E_{\phi_2} \tag{6.4.33}$$

and inserting the (renormalized) solution above into (6.4.9) we finally find

$$A_{\phi\chi\chi}(t) = \frac{-iM^*_{\pi\to\phi_1\phi_2} M^*_{\phi_1\to\chi_1\chi_2}}{E^\pi - E_{\phi\phi} - \frac{i}{2}(\Gamma_\pi - \Gamma_{\phi_1})} \int_0^t dt' \left[e^{-i(E^\pi - E_{\phi\chi\chi} - i\frac{\Gamma_\pi}{2})t'} - e^{-i(E_{\phi\phi} - E_{\phi\chi\chi} - i\frac{\Gamma_{\phi_1}}{2})t'} \right]$$
$$= \frac{M^*_{\pi\to\phi_1\phi_2} M^*_{\phi_1\to\chi_1\chi_2}}{E^\pi - E_{\phi\phi} - \frac{i}{2}(\Gamma_\pi - \Gamma_{\phi_1})} \left[\frac{e^{-i(E^\pi - E_{\phi\chi\chi} - i\frac{\Gamma_\pi}{2})t} - 1}{E^\pi - E_{\phi\chi\chi} - i\frac{\Gamma_\pi}{2}} - \frac{e^{-i(E_{\phi\phi} - E_{\phi\chi\chi} - i\frac{\Gamma_{\phi_1}}{2})t} - 1}{E_{\phi\phi} - E_{\phi\chi\chi} - i\frac{\Gamma_{\phi_1}}{2}} \right] \tag{6.4.34}$$

Passing back to the Schrodinger picture and after renormalization of the single particle energies it follows that

$$A_\pi(t) \to e^{-iE^\pi\,t}\, A_\pi(t) = C_\pi(t) \tag{6.4.35}$$
$$A_{\phi\phi}(t) \to e^{-iE_{\phi\phi}\,t}\, A_{\phi\phi}(t) = C_{\phi\phi}(t) \tag{6.4.36}$$
$$A_{\phi\chi\chi}(t) \to e^{-iE_{\phi\chi\chi}\,t}\, A_{\phi\chi\chi}(t) = C_{\phi\chi\chi}(t)\,, \tag{6.4.37}$$

finally matching the results (6.2.44,6.2.48, 6.2.51) that were obtained via Laplace transform.

Discussion: The analysis above shows that the field theoretical generalization of the Wigner-Weisskopf method provides a real time realization of the non-perturbative resummation akin to the Dyson resummation of self-energies in the propagators and yields a resummation of secular terms that grow in time. The Markovian approximation based on a derivative expansion that relies on a separation of time scales valid in the weak coupling

regime in the long time limit is akin to the Breit-Wigner or narrow width approximation in that it captures reliably the decay of resonances. Furthermore, the equivalence with the results of the previous section, clearly proves that the Wigner Weisskopf method is manifestly unitary.

While both methods, are equivalent in *Minkowski space time*, the quantum field theoretical Wigner-Weisskopf method features the distinct advantage of direct applicability in the cosmological context wherein the expansion implies an explicitly time dependent interaction Hamiltonian *in the Schroedinger picture* as a consequence of the cosmological expansion. To lowest order this method has been applied in cosmology in refs.[175, 231], however the purpose of this work is to extend it to the hitherto unexplored case of *radiative cascade decay*. The results of this section indicate the reliability of the method thereby bolstering the case for its implementation in cosmology, which will be the subject of future study.

6.5 POSSIBLE PHENOMENOLOGICAL CONSEQUENCES AND CORRELATIONS.

6.5.1 Possible phenomenological consequences:

A *possible* phenomenological consequence may emerge if heavy sterile neutrinos ν_s exist and mix with active neutrinos. For example, consider the case of $\nu_\mu - \nu_s; \nu_e - \nu_s$ mixing in π-decay, if ν_s features a mass[1] $2m_e < m_s < m_\pi - m_\mu$ (here we neglect the mass of the "active" neutrino mass eigenstates) then the intermediate state with a ν_s becomes resonant and can decay either via charged or neutral current interactions into $e^+e^-\nu_e$. This process is depicted in the Fermi limit in fig. (33) and would correspond to an *appearance* contribution.

If the lifetime of the heavy sterile neutrino is very long for example of the order of the baseline in long baseline neutrino experiments, the process above will yield a contribution to the appearance probability. If on the other hand the lifetime of the putative sterile neutrino is short, then it can decay on distances shorter than the oscillation length of active

[1] We here simply refer to ν_s as the sterile-like *mass eigenstate*.

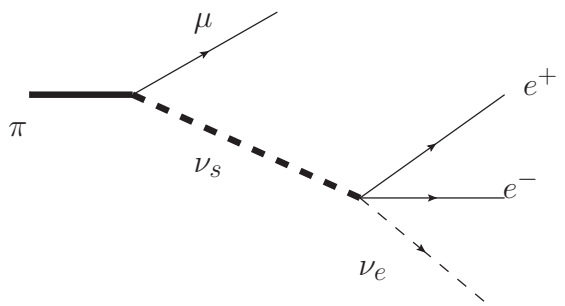


Figure 33: A *possible* process: $\pi \to \mu\nu_s \to e^+e^-\,\nu_e$.

neutrinos and this process would contribute to the appearance probability with oscillations and the concomitant distortion in the energy spectrum, but on length scales shorter than the oscillation length. Clearly such process will be suppressed by a product $U_{es}U_{\mu s}$. Furthermore unlike oscillations of active neutrinos, the time dependence in this case would be damped exponentially on a scale of the order of the decay length of the sterile neutrino. For example if the heavy sterile neutrino decays via a charged current vertex, and neglecting the electron mass

$$\Gamma_{\nu_s} \simeq \frac{G_F^2\, m_{\nu_s}^5}{192\pi^3}\, |U_{es}|^2 \tag{6.5.1}$$

with a decay length

$$c\tau_{\nu_s} \simeq \frac{700}{|U_{es}|^2}\,\text{mts}\,. \tag{6.5.2}$$

For $|U_{es}| \ll 1$ the maximum in the probability of the intermediate state with the sterile neutrino resonance occurs at a distance

$$ct^* \sim -10\ln|U_{es}|\,\text{mts} \tag{6.5.3}$$

this is approximately the distance away from the decay region of the parent pion at which the final state leptons are produced, namely the "decay vertex". A detection of the charge lepton a distance L away from the production region *may* mis-identify these charged leptons as being produced by neutrinos resulting from the decay of the parent meson or from oscillations

between active neutrinos, the displaced decay vertex of the heavy sterile neutrino would then imply a *shortening* the effective baseline by a large factor if $|U_{es}| \ll 1$.

Similar "rare pion decay" processes have been discussed within the context of lepton flavor violation in ref.[222] but without addressing the full time evolution.

6.5.2 Entanglement and correlations.

The asymptotic state after the parent particle and the intermediate resonant state have decayed is given by

$$|\Psi(\infty)\rangle = \sum_{\vec{q},\vec{p}} C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};\infty)\left|\phi_{2,\vec{k}-\vec{q}}\chi_{1,\vec{p}}\chi_{2,\vec{q}-\vec{p}}\right\rangle , \tag{6.5.4}$$

where

$$C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};\infty) = \frac{M^{*}_{\pi\to\phi_1\phi_2}\, M^{*}_{\phi_1\to\chi_1\chi_2}\, e^{-iE_{\phi\chi\chi}\,t}}{\left[(E_{\phi\chi\chi}-E^{\pi})+i\frac{\Gamma_\pi}{2}\right]\left[(E_{\phi\chi\chi}-E_{\phi\phi})+i\frac{\Gamma_{\phi_1}}{2}\right]} \tag{6.5.5}$$

this is entangled state of three particles, which apart from $|\phi_{2,\vec{k}-\vec{q}}\rangle$ is very similar to the asymptotic entangled two-photon state from a radiative cascade of two-level atomic systems in quantum optics[204, 232].

Extrapolating the results to the case of phenomenological relevance discussed above, the asymptotic state $|\Psi(\infty)\rangle$ is an entangled state of four particles. Quantum entanglement entails correlations, these are completely determined by the amplitudes $C(\infty)$, which depend on the mass and width of the resonant state and the matrix elements. These correlations will be manifest as intensity Hanbury-Brown-Twiss correlations among the charged lepton pairs produced by the decay of the resonant state that *could* reveal important information on the properties of heavy sterile neutrinos in the intermediate state.

We postpone a deeper study of the time dependence of these phenomenological consequences including mixing and oscillations to a future article.

6.6 CONCLUSIONS AND FURTHER QUESTIONS:

Cascade decays via resonant intermediate states are of interdsiciplinary interest, being ubiquitous in particle physics and quantum optics, and more recently are discussed within the context of inflationary cosmology. In this article we generalize methods of quantum optics to the realm of quantum field theory to study the real time dynamics of cascade decay in a model quantum field theory of generic fields, however the conclusions are general.

The method is based on a hierarchical solution of the coupled equations for the amplitudes of the initial, intermediate and final multiparticle states. We show that a solution of the equations up to a given order in the interaction yields a non-perturbative resummation *a la Dyson* in terms of the self-energies of the initial and intermediate resonant states. We analyze the time evolution of the amplitudes and probabilities and show that unitary time evolution is manifest as a probability "flow" from the initial through the intermediate resonant and to the final state. When the decay width of the initial parent particle Γ_π is much larger than that of the intermediate resonant state Γ_{ϕ_1} there is a "bottleneck" in the evolution, the probability of the intermediate resonant state grows to a maximum nearly saturating unitarity on a time scale given

$$t^* = \frac{\ln\left[\frac{\Gamma_\pi}{\Gamma_{\phi_1}}\right]}{\Gamma_\pi - \Gamma_{\phi_1}} \tag{6.6.1}$$

and decays on a longer time scale $\simeq 1/\Gamma_{\phi_1}$, whereas in the opposite limit the population of the resonant state does not build up substantially and the probability flows almost directly from the initial to the final state on a time scale $1/\Gamma_\pi$.

We provided an alternative formulation in terms of a quantum field theoretical generalization of the Wigner-Weisskopf method in quantum optics. This method provides a non-perturbative resummation of secular terms in time and reliably describes the time evolution of intermediate resonant states. While both methods are equivalent, the quantum field theoretical generalization of Wigner-Weisskopf is directly applicable in the cosmological setting where the interaction Hamiltonian is explicitly dependent on time.

We *conjecture* on potential phenomenological implications, in particular in the case of pseduscalar meson decay via possible heavy sterile neutrinos as intermediate resonant states.

Their production and decay into "active" neutrinos and charged leptons may have experimental relevance, we argue that cascade decay in this case may lead to "displaced decay vertices" which may result in important corrections to the baseline dependence of the appearance (and dissapearance) probabilities. Asymptotically at long times after both the parent and resonant intermediate particles decay, the final state is a quantum entangled many particle state that features quantum correlations which are completely determined by the asymptotic amplitudes. We also conjecture that quantum entanglement of the final state may translate in intensity Hanbury-Brown-Twiss correlations that may reveal information on the mass and width of the intermediate resonant state.

The analysis and results presented above may be important in resonant leptogenesis[233, 234, 235, 236], which is typically studied by implementing powerful non-equilibrium methods such as the Kadanoff-Baym and Keldysh formulations to obtain the kinetic description of the distribution functions that include resonant cross sections. However, the application of the results obtained in this article to the important case of resonant leptogenesis is not direct: we obtained the evolution equations for the amplitudes of the many particle state resulting from the decay of an initial *single particle state*. Instead the Kadanoff-Baym and Keldysh approaches to resonant leptogenesis focus on the time evolution of *distribution functions* which are ensemble averages in a non-equilibrium *density matrix*. In order to apply the methods and results obtained above to this important case, first we must understand how to implement the non-perturbative methods described above to the case of a non-equilibrium density matrix rather than a single particle initial state. The importance of resonant leptogenesis motivates further study to generalize the results obtained here to the case of non-equilibrium ensembles, correlation and distribution functions on which we expect to report in the future.

Further questions:

In this article we focused on the description of the time evolution of cascade decay in a simple scenario with only one intermediate resonant state. An important case that remains to be studied is that of several intermediate resonant states that may result from mixing. In this case there will emerge interference phenomena manifest as oscillations, and if the

intermediate states are nearly degenerate these oscillations and interference may lead to important dynamics at long times. This is the case for meson mixing relevant for CP and CPT violations. As discussed in the introduction cascade "mixing" is typically studied as sequential events in (proper) time, however the analysis presented in the previous sections suggests that there may be important corrections from the time dependence of the amplitudes that *may* prove to be relevant to the experimental analysis. These questions along with an assessment of potential impact on neutrino oscillation experiments from a resonant heavy sterile neutrino, and the study of correlations in the final state as potential indicators of properties of the intermediate resonant state merit further study, which is postponed to a future article.

6.7 APPENDICES

6.7.1 Analysis of $C_{\phi\chi\chi}(\vec{k},\vec{q},\vec{p};t)$

To make the notation in this appendix more compact we introduce the following variables

$$E_{\phi\chi\chi}-E^{\pi}_{k}=\eta \;\;;\;\; E_{\phi\phi}-E^{\pi}_{k}=\sigma \;\;;\;\; \Delta=\Gamma_{\pi}-\Gamma_{\phi_1} \;\;;\;\; \Sigma=\Gamma_{\pi}+\Gamma_{\phi_1} \tag{6.7.1}$$

and suppressing the momenta labels, we write

$$C_{\phi\chi\chi}(t)\equiv D(t)+C_{\phi\chi\chi}(\infty)\,. \tag{6.7.2}$$

The contribution from the term $|C_{\phi\chi\chi}(\infty)|^2$ has been analyzed above, leading to eqn. (6.3.14), for the remaining terms we find

$$|D(t)|^2+D(t)C^{*}_{\phi\chi\chi}(\infty)+D^{*}(t)C_{\phi\chi\chi}(\infty)=|M_{\pi\to\phi_1\phi_2}|^2\;\;|M_{\phi_1\to\chi_1\chi_2}|^2\Big[(a)+(b)+(c)+(d)+(e)\Big] \tag{6.7.3}$$

$$(a)=\frac{1}{\sigma^2+\left(\frac{\Delta}{2}\right)^2}\left[\frac{e^{-\Gamma_\pi t}-e^{-\frac{\Gamma_\pi}{2}t}(e^{i\eta t}+e^{-i\eta t})}{\eta^2+\left(\frac{\Gamma_\pi}{2}\right)^2}\right] \tag{6.7.4}$$

$$(b) = \frac{1}{\sigma^2 + \left(\frac{\Delta}{2}\right)^2}\left[\frac{e^{-\Gamma_{\phi_1}\,t} - e^{-\frac{\Gamma_{\phi_1}}{2}\,t}(e^{i(\eta-\sigma)\,t} + e^{-i(\eta-\sigma)\,t})}{(\eta-\sigma)^2 + \left(\frac{\Gamma_{\phi_1}}{2}\right)^2}\right] \tag{6.7.5}$$

$$(c) = \frac{1}{\sigma^2 + \left(\frac{\Delta}{2}\right)^2}\left[\frac{e^{i\eta\,t}\,e^{-\frac{\Gamma_\pi}{2}\,t} - e^{i\sigma\,t}\,e^{-\frac{\Sigma}{2}\,t}}{\left(\eta + i\frac{\Gamma_\pi}{2}\right)\left(\eta - \sigma - i\frac{\Gamma_{\phi_1}}{2}\right)}\right] \tag{6.7.6}$$

$$(d) = \frac{1}{\sigma^2 + \left(\frac{\Delta}{2}\right)^2}\left[\frac{e^{i(\eta-\sigma)\,t}\,e^{-\frac{\Gamma_{\phi_1}}{2}\,t} - e^{-i\sigma\,t}\,e^{-\frac{\Sigma}{2}\,t}}{\left(\eta - i\frac{\Gamma_\pi}{2}\right)\left(\eta - \sigma + i\frac{\Gamma_{\phi_1}}{2}\right)}\right] \tag{6.7.7}$$

$$(e) = \frac{1}{\sigma^2 + \left(\frac{\Delta}{2}\right)^2}\left[\frac{e^{-i\eta\,t}\,e^{-\frac{\Gamma_\pi}{2}\,t}}{\left(\eta - i\frac{\Gamma_\pi}{2}\right)\left(\eta - \sigma + i\frac{\Gamma_{\phi_1}}{2}\right)} + \frac{e^{-i(\eta-\sigma)\,t}\,e^{-\frac{\Gamma_{\phi_1}}{2}\,t}}{\left(\eta + i\frac{\Gamma_\pi}{2}\right)\left(\eta - \sigma - i\frac{\Gamma_{\phi_1}}{2}\right)}\right]. \tag{6.7.8}$$

The resonant denominators result in that the dominant contribution in the narrow widths limit are proportional to $\delta(\sigma)\delta(\eta)$, in order to extract the proportionality factors we integrate the above expressions in the complex σ, η planes where the resonant denominators yield complex poles. We find

$$(a) + (b) = -\frac{(2\pi)^2}{|\Delta|}\left\{\frac{e^{-\Gamma_\pi\,t}}{\Gamma_\pi}\;\delta(\eta) + \frac{e^{-\Gamma_{\phi_1}\,t}}{\Gamma_{\phi_1}}\;\delta(\eta-\sigma)\right\}\;\delta(\sigma)\,. \tag{6.7.9}$$

The integrals in the complex planes of $(c), (d)$ feature vanishing residues at the complex poles in $\eta, \eta - \sigma$, therefore these integrals yield subleading contributions in the narrow width limit. Finally by the same procedure we find

$$(e) = 2\frac{(2\pi)^2}{|\Delta|}\;\;\frac{e^{-\frac{|\Delta|}{2}\,t}\;e^{-\frac{\Sigma}{2}\,t}}{\frac{1}{2}\,(|\Delta| + \Sigma)}\;\delta(\eta)\;\delta(\sigma)\,. \tag{6.7.10}$$

The final result is given by

$$(a)+(b)+(c)+(d)+(e) = \frac{(2\pi)^2}{\Gamma_\pi - \Gamma_{\phi_1}}\left\{\frac{e^{-\Gamma_\pi\,t}}{\Gamma_\pi} - \frac{e^{-\Gamma_{\phi_1}\,t}}{\Gamma_{\phi_1}}\right\}\delta(E^\pi_k - E_{\phi\chi\chi})\;\delta(E^\pi_k - E_{\phi\phi}) \tag{6.7.11}$$

6.7.2 The long time limit of eqn. (6.4.19)

Introducing

$$\rho(\omega) = \sum_{\vec{q}} |M_{\phi_1 \to \chi_1 \chi_2}|^2 \delta(\omega - E_{\phi\chi\chi}) \,, \tag{6.7.12}$$

we can write

$$\begin{aligned} \int_0^t dt' W_0^\phi(t', t') &= i\, t \int_{-\infty}^{\infty} d\omega' \, \frac{\rho(\omega')}{(E_{\phi\phi} - \omega')} \left[1 - \frac{\sin(\omega' - E_{\phi\phi})t}{(\omega' - E_{\phi\phi})t} \right] \\ &+ \int_{-\infty}^{\infty} d\omega' \, \frac{\rho(\omega')}{(E_{\phi\phi} - \omega')^2} \left[1 - \cos\left[(\omega' - E_{\phi\phi})t\right] \right] . \end{aligned} \tag{6.7.13}$$

Asymptotically as $t \to \infty$, these integrals approach:

$$\int_{-\infty}^{\infty} d\omega' \, \frac{\rho(\omega')}{(E_{\phi\phi} - \omega')} \left[1 - \frac{\sin(\omega' - E_{\phi\phi})\, t}{(\omega' - E_{\phi\phi})\, t} \right] \xrightarrow[t \to \infty]{} \mathcal{P} \int_{-\infty}^{\infty} d\omega' \, \frac{\rho(\omega')}{(E_{\phi\phi} - \omega')} \tag{6.7.14}$$

$$\int_{-\infty}^{\infty} d\omega' \, \frac{\rho(\omega')}{(E_{\phi\phi} - \omega')^2} \left[1 - \cos\left[(\omega' - E_{\phi\phi})t\right] \right] \xrightarrow[t \to \infty]{} \pi\, t\, \rho(E_{\phi\phi}) \,. \tag{6.7.15}$$

The second integral above can be easily recognized as the usual Fermi's Golden rule by taking the time derivative .

7.0 PRE-SLOW ROLL INITIAL CONDITIONS: INFRARED ASPECTS AND LARGE SCALE POWER SUPPRESSION DURING INFLATION.

Based on: (ref. [311])

L. Lello, D. Boyanovsky, R. Holman, Phys. Rev. D 89, 063533 (2014)

7.1 INTRODUCTION

Inflation provides a solution to the horizon and flatness problems and a mechanism for generating scalar (curvature) and tensor (gravitational wave) quantum fluctuations[237, 238, 239, 240] which seed the small temperature inhomogeneities in the CMB upon reentering the particle horizon during recombination. Although there are several different inflationary scenarios most of them predict a nearly gaussian and nearly scale invariant power spectrum of adiabatic fluctuations.For reviews see[241, 245, 242, 243, 244].

Observations of the cosmic microwave background (CMB) offer compelling evidence in support of the inflationary paradigm, confirming that anisotropies are well described by adiabatic, gaussian and nearly scale invariant fluctuations[246, 210, 63] and are beginning to discriminate among different scenarios.

Recent results from the Planck collaboration[63] have provided the most precise analysis of the (CMB) to date, confirming the main features of the inflationary paradigm, but at the same time highlighting perplexing large scale anomalies, some of them, such as a low quadrupole, dating back to the early observations of the Cosmic Background Explorer

(COBE)[247, 248], confirmed with greater accuracy by WMAP[249] and Planck[63]. The most recent Planck[63] data still finds a statistically significant discrepancy at low multipoles, reporting a power deficit $5-10\%$ at $l \lesssim 40$ with $2.5-3\,\sigma$ significance. This puzzling and persistent result stands out in an otherwise consistent picture of ΛCDM insofar as the (CMB) power spectrum is concerned.

The interpretation and statistical significance of these anomalies is a matter of much debate, but being associated with the largest scales, hence the most primordial aspects of the power spectrum, their observational evidence is not completely dismissed. The possible origin of the large scale anomalies is vigorously discussed, whether these are of primordial origin or a consequence of the statistical analysis (masking) or secondary anisotropies is still an open question. Recent studies claim the removal of some of the large scale anomalies (including the suppression of power of the low multipoles) after substraction of the integrated Sachs-Wolfe effect[250, 251], however a different analysis of the WMAP9[210] data still finds a statistically significant discrepancy at low multipoles[211], suggesting that the possibility of the primordial origin of the large scale anomalies merits further study. Recent analysis of this lack of power at low l[211] and large angles[252], suggests that while limited by cosmic variance, the possibility of the primordial origin of the large scale anomalies cannot be dismissed and merits further study.

The simpler inflationary paradigm that successfully explains the cosmological data relies on the dynamics of a scalar field, the inflaton, evolving slowly during the inflationary stage with the dynamics determined by a fairly flat potential. This simple, yet observationally supported inflationary scenario is referred to as slow-roll inflation[245, 241, 242, 243, 244]. Within this scenario wave vectors of cosmological relevance cross the Hubble radius during inflation with nearly constant amplitude leading to a nearly scale invariant power spectrum. The quantization of the gaussian fluctuations (curvature and tensor) is carried out by imposing a set of initial conditions so that fluctuations with wavevectors deep inside the Hubble radius are described by Minkowski space-time free field mode functions. These are known as Bunch-Davies initial conditions[253] (see for example[245, 241, 243, 244] and references therein).

The issue of modifications of these initial conditions and the potential impact on the inflationary power spectra[254, 255, 256, 257, 258, 259, 260, 261, 262, 208, 263, 264], enhancements to non-gaussianity[265, 266, 267, 268, 269, 270, 271], and large scale structure[272] have been discussed in the literature. Whereas the recent results from Planck[63] provide tight constraints on primordial non-gaussianities including modifications from initial conditions, these constraints *per se* do not apply directly to the issue of initial conditions on other observational aspects.

Non-Bunch-Davies initial conditions arising from a pre-slow roll stage during which the (single) inflaton field features a "fast-roll" dynamics have been proposed as a *possible* explanation of power suppression at large scales[273, 274, 275, 276, 277]. Alternative pre-slow-roll descriptions in terms of interpolating scale factors pre (and post) inflation have been discussed in ref.[278] and the impact of initial conditions from high energy models on power spectra and non-gaussianities and the tensor to scalar ratio was studied in ref.[279, 280, 281, 282].

More recently a detailed analysis of modifications of power spectra for curvature and tensor perturbations from a kinetic dominated pre-slow roll stage has been reported[283].

The largest scales that manifest the suppression in the power spectrum correspond to fluctuations whose wavevectors exited the Hubble radius about 60-e-folds before the end of inflation, therefore *if* the large scale anomalies are of primordial origin and herald new physics, an explanation must be sought in the *infrared* sector of inflationary perturbations.

It has been recognized that the contribution from super-Hubble fluctuations of massless (or nearly massless) fields in de Sitter (or nearly de Sitter) inflation to loop corrections of cosmological correlation functions lead to infrared and secular divergences that hinder the reliability of the perturbative expansion[284, 285, 286, 287, 288, 289, 183, 290]. These divergences invalidate the semiclassical approximation[199] and require non-perturbative resummations[291, 292, 175, 293] or kinetic[192] treatments.

In the seminal work of ref.[294] it was shown that resummation of infrared and secular divergences leads to the dynamical generation of mass, a result that was further explored in ref.[295] and more recently a self-consistent mechanism of mass generation for scalar fields through infrared fluctuations has been suggested[199, 200, 296, 297, 290, 298, 175].

Furthermore, the lack of a global time-like killing vector leads to remarkable physical effects in inflationary cosmology, for example it implies the lack of kinematic thresholds (a direct consequence of energy-momentum conservation) and the decay of fields even in their own quanta[190, 229, 175] a result that was also investigated for massive fields in ref.[191, 201] and confirmed in general in ref.[194].

If a parent particle decays into two or more daughter particles, the quantum state that describes the daughter particles is an *entangled state*[198], the entanglement is a consequence of conservation laws, such as momentum, angular momentum etc. Recently it was recognized that in inflationary cosmology the decay of a particle into sub and superhorizon quanta produces an *entangled state* with quantum correlations across the Hubble radius[299].

Motivations, goals and summary of results: Our study is motivated by the persistency of the power suppression at large scales as evidenced in the Planck data[63] and the possibility that these anomalies are of primordial origin and reflect novel physical infrared effects with observational consequences.

Our goals in this article are two-fold: i) to study in detail the modifications of initial conditions within the paradigm of single field inflation but described by an early, pre-slow-roll stage in which the inflaton field undergoes "fast-roll" dynamics as proposed in refs.[273, 274, 275, 276, 277], ii) to assess how the modified initial conditions impact infrared phenomena in scalar fields. In particular we focus on the impact of non-Bunch-Davies initial conditions as a consequence of a pre-slow-roll stage on non-perturbative phenomena, such as dynamical mass generation, decay of quanta and superhorizon correlations arising from the quantum entanglement of the daughter particles. To the best of our knowledge, the effect of initial conditions on infrared effects such as dynamical mass generation and decay of single particle excitations has not been studied.

We consider the case in which non-Bunch-Davies initial conditions during inflation are a consequence of a pre-slow roll stage during which the inflaton undergoes fast roll dynamics. This "fast-roll" stage prior to slow roll results in a potential in the equations of motion of gaussian fluctuations and lead to a change of initial conditions during the slow roll stage via Bogoliubov coefficients.

We begin with a description of a fast-roll stage dominated by the kinetic term of the inflaton and follow with a detailed analysis of superhorizon and subhorizon behavior of the Bogoliubov coefficients describing non-Bunch-Davies initial conditions during the inflationary stage and how these modify the large scale power spectrum of fluctuations. The effect of these non-Bunch-Davies initial conditions is encoded in the power spectra of scalar and tensor perturbations via a *transfer function* $\mathcal{T}(k)$.

Implementing methods from the quantum theory of scattering, we provide an analytic study of the superhorizon and subhorizon limits of the initial condition transfer function $\mathcal{T}(k)$ and find that for superhorizon momenta

$$\mathcal{T}(k) \simeq \mathcal{T}(0) + \mathcal{O}(k^2)$$

and obtain an explicit expression for $\mathcal{T}(0)$. For subhorizon momenta we find

$$\mathcal{T}(k) \simeq 1 + \mathcal{O}(1/k^4)\,.$$

We extract the form of the mode functions modified by these initial conditions in the superhorizon limit and study in detail how this transfer function modifies the infrared behavior in typical scalar field theories, in particular the modification of dynamically generated masses and the width of the single particle states.

We find that the dynamically generated masses induced by these infrared divergences depend non-analytically on the transfer function. As a consequence of dynamical mass generation the scalar power spectrum features anomalous dimensions that depend non-analytically on $\mathcal{T}(0)$. The decay width of single particle quanta and the entanglement entropy from integrating out superhorizon fluctuations depend also on this quantity. We find that when the potential produced by the fast rolling inflaton is attractive $\mathcal{T}(0) < 1$ and the power spectra are suppressed at large scales. This suppression is also manifest in the dynamically generated masses, anomalous dimensions, decay widths and entanglement entropy.

7.2 FAST ROLL STAGE:

In this section we summarize the main aspects of a kinetic dominated pre-slow roll stage or "fast-roll" stage. More details and a complete analysis of the matching to slow roll can be found in ref.[283].

We consider a spatially flat Friedmann-Robertson-Walker (FRW) cosmology with

$$ds^2 = dt^2 - a^2(t)(d\vec{x})^2 = C^2(\eta)[d\eta^2 - (d\vec{x})^2] \;\; ; \;\; C(\eta) \equiv a(t(\eta)) \, , \tag{7.2.1}$$

where t and η stand for cosmic and conformal time respectively and consider curvature and tensor perturbations. The dynamics of the scale factor in single field inflation is determined by Friedmann and covariant conservation equations

$$H^2 = \frac{1}{3M_{Pl}^2}\left[\frac{1}{2}\dot{\Phi}^2 + V(\Phi)\right] \;\; ; \;\; \ddot{\Phi} + 3H\dot{\Phi} + V'(\Phi) = 0 \, . \tag{7.2.2}$$

During the slow roll near de Sitter stage,

$$H_{sr}^2 \simeq \frac{V_{sr}(\Phi)}{3M_{Pl}^2} \;\; ; \;\; 3H\dot{\Phi} + V'_{sr}(\Phi) \simeq 0 \, . \tag{7.2.3}$$

This stage is characterized by the smallness of the (potential) slow roll parameters[241, 245, 242, 243, 244]

$$\epsilon_V = \frac{M_{Pl}^2}{2}\left[\frac{V'_{sr}(\Phi)}{V_{sr}(\Phi)}\right]^2 \simeq \frac{\dot{\Phi}_{sr}^2}{2M_{Pl}^2H^2} \quad , \quad \eta_V = M_{Pl}^2\frac{V''_{sr}(\Phi)}{V_{sr}(\Phi)} \, , \tag{7.2.4}$$

(here $M_{Pl} = 1/\sqrt{8\pi\, G}$ is the *reduced* Planck mass). The potential slow roll parameters ϵ_V, η_V which have been constrained by Planck and WMAP-polarization (Planck+WP)[63] to be $\epsilon_V < 0.008 \; (95\%\, CL); \eta_V = -0.010^{+0.005}_{-0.011}$.

Instead, in this section we consider an initial stage dominated by the kinetic term, namely a fast roll stage, thereby neglecting the term V' in the equation of motion for the inflaton, (7.2.2) and consider the potential to be (nearly) constant and equal to the potential during

the slow roll stage, namely $V(\Phi) \simeq V(\Phi_{sr}) \equiv V_{sr}$. In the following section we relax this condition in a consistent expansion in $\sqrt{\epsilon_V}$.

$$H^2 = \Big(\frac{\dot{a}}{a}\Big)^2 = \frac{1}{3M_{Pl}^2}\left[\frac{1}{2}\dot{\Phi}^2 + V_{sr}\right] \tag{7.2.5}$$

$$\ddot{\Phi} + 3H\dot{\Phi} \simeq 0\,. \tag{7.2.6}$$

The solution to (7.2.6) is given by

$$\dot{\Phi}(t) = \dot{\Phi}_i\Big(\frac{a_i}{a(t)}\Big)^3\,, \tag{7.2.7}$$

an initial value of the velocity damps out and the slow roll stage begins when $\ddot{\Phi} \ll 3H_{sr}\dot{\Phi} \simeq -V'_{sr}(\Phi)$. During the slow roll stage when $3H_{sr}\dot{\Phi}_{sr} \simeq -V'_{sr}$ it follows that

$$\frac{3\dot{\Phi}_{sr}^2}{2V_{sr}} = \epsilon_V\,. \tag{7.2.8}$$

The dynamics enters the slow roll stage when $\dot{\Phi} \sim \mathcal{O}(\sqrt{\epsilon_V})$ as seen by (7.2.4). To a first approximation, we will assume that Eq.(7.2.7) holds not only for the kinetically dominated epoch, but also until the beginning of slow roll ($\dot{\Phi}^2 \sim \epsilon_V$). The dynamics enters the slow roll stage at a value of the scale factor $a(t_{sr}) \equiv a_{sr}$ so that

$$\dot{\Phi}_{sr}a_{sr}^3 = \dot{\Phi}_i a_i^3\,. \tag{7.2.9}$$

We now use the freedom to rescale the scale factor to set

$$a(t_{sr}) = a_{sr} = 1\,, \tag{7.2.10}$$

this normalization is particularly convenient to establish when a particular mode crosses the Hubble radius during slow roll, an important assessment in the analysis below.

In terms of these definitions and eqn. (7.2.9), we have that during the fast roll stage

$$\dot{\Phi}(t) = \frac{\dot{\Phi}_{sr}}{a^3(t)}\,. \tag{7.2.11}$$

Introducing

$$H_{sr}^2 \equiv \frac{V_{sr}}{3M_{Pl}^2}\,, \tag{7.2.12}$$

Friedmann's equation becomes

$$\frac{\dot{a}(t)}{a(t)} = H_{sr}\left[1 + \frac{\epsilon_V}{3\,a^6(t)}\right]^{1/2} . \tag{7.2.13}$$

This equation for the scale factor can be readily integrated to yield the solution

$$a(t) = \left[\left(\frac{\epsilon_V}{3}\right)^{1/2} \sinh[\theta(t)]\right]^{1/3} \quad ; \quad \theta(t) = \theta_0 + 3H_{sr}t \tag{7.2.14}$$

where θ_0 is an integration constant chosen to be

$$e^{-\theta_0} = \sqrt{\frac{\epsilon_V}{12}}\,, \tag{7.2.15}$$

so that at long time $a(t) = e^{H_{sr}t}$. The slow roll stage begins when $a(t_{sr}) = 1$ which corresponds to the value of $\theta_{sr} = \theta(t_{sr})$ given by

$$e^{-\theta_{sr}} = f(\frac{\epsilon_V}{3}) \tag{7.2.16}$$

where to simplify notation later we defined

$$f(s) = \frac{\sqrt{s}}{1 + \sqrt{1+s}}\,. \tag{7.2.17}$$

Introducing the dimensionless ratio of kinetic to potential contributions at the initial time t_i

$$\frac{\dot{\Phi}_i^2}{2V_{sr}} = \kappa\,, \tag{7.2.18}$$

and *assuming* that the potential does not vary very much between the initial time and the onset of slow roll (this is quantified below), it follows from (7.2.9) that

$$a_i^6 = \frac{\dot{\Phi}_{sr}^2}{2V_{sr}\kappa} = \frac{\epsilon_V}{3\kappa} \tag{7.2.19}$$

where we have used (7.2.8). Combining this result with (7.2.14), we find that at the initial time $\theta_i = \theta(t_i)$ is given by

$$e^{-\theta_i} = f(\kappa)\,. \tag{7.2.20}$$

Let us introduce

$$\varepsilon(t) = -\frac{\dot{H}}{H^2} = \frac{\dot{\Phi}^2}{2M_{Pl}^2H^2} = \frac{\epsilon_V}{a^6(t) + \frac{\epsilon_V}{3}} \tag{7.2.21}$$

where we have used the results (7.2.7,7.2.8,7.2.9) from which it is clear that for $\epsilon_V \ll 1$ the slow roll stage begins at $a = 1$ when $\varepsilon = \epsilon_V + \mathcal{O}(\epsilon_V^2)$. With $a(t)$ given by (7.2.14), it follows that

$$\varepsilon(t) = \frac{3}{\cosh^2[\theta(t)]}\,, \tag{7.2.22}$$

therefore $0 \leq \varepsilon \leq 3$, and

$$H(t) = \frac{H_{sr}}{\tanh[\theta(t)]}\,. \tag{7.2.23}$$

The acceleration equation written in terms of $\varepsilon(t)$ is given by

$$\frac{\ddot{a}}{a} = H^2(t)(1 - \varepsilon(t))\,, \tag{7.2.24}$$

so that the inflationary stage begins when $\varepsilon(t) = 1$. At the initial time

$$\varepsilon(t_i) = \frac{3\kappa}{1+\kappa} \tag{7.2.25}$$

hence, for $\kappa > 1/2$ the early stage of expansion is deccelerated and inflation begins when $\varepsilon(t_{inf}) = 1$.

It proves convenient to introduce the variable

$$x(t) = e^{-\theta(t)/3} = \left[\frac{\epsilon_V}{12}\right]^{1/6} e^{-H_{sr}t}\,, \tag{7.2.26}$$

with

$$x_i \equiv x(t_i) = [f(\kappa)]^{1/3} \quad ; \quad x_{sr} \equiv x(t_{sr}) = [f(\epsilon_V/3)]^{1/3}\,. \tag{7.2.27}$$

where $f(s)$ is given by eqn. (7.2.17), and write a, H, ε in terms of this variable leading to

$$a(x) = \left[\frac{\epsilon_V}{12}\right]^{1/6} \frac{\left[1 - x^6\right]^{1/3}}{x}\,, \tag{7.2.28}$$

$$H(x) = H_{sr}\frac{\left[1 + x^6\right]}{\left[1 - x^6\right]}\,, \tag{7.2.29}$$

$$\varepsilon(x) = \frac{12\,x^6}{\left[1 + x^6\right]^2}\,. \tag{7.2.30}$$

,

Conformal time $\eta(t)$ defined to vanish as $t \to \infty$ is given by

$$\begin{aligned} \eta(t) &= \int_{\infty}^{t} \frac{dt'}{a(t')} = \int_{\infty}^{a(t)} \frac{da}{a^2\, H(a)} \\ &= -\frac{1}{a(t)H(t)} + \int_{\infty}^{t} \varepsilon(t') \frac{dt'}{a(t')} \end{aligned} \tag{7.2.31}$$

where we integrated by parts and used the definition of ε given by eqn. (7.2.21). Adding and subtracting ϵ_V we find

$$\eta(t) = -\frac{1}{a(t)H(t)(1-\epsilon_V)} + \frac{\epsilon_V}{(1-\epsilon_V)} \int_{\infty}^{t} \Big[\frac{\varepsilon(t')}{\varepsilon_V} - 1\Big] \frac{dt'}{a(t')}\,, \tag{7.2.32}$$

The argument of the integrand in the second term in (7.2.32) vanishes to leading order in ϵ_V, η_V in the slow roll phase (when $t > t_{sr}$). Therefore, during slow roll, $\eta = -1/aH(1-\epsilon_V)$. It will be convenient to write η in terms of the variable x (7.2.26), it is given by

$$\begin{aligned} \eta(x) &= -\frac{1}{H_{sr}\,(1-\epsilon_V)} \Big(\frac{12}{\epsilon_V}\Big)^{1/6} \Bigg\{ \frac{x(1-x^6)^{2/3}}{(1+x^6)} \\ &+ \epsilon_V \int_{x_{sr}}^{x} \frac{dy}{[1-y^6]^{1/3}} \left[\frac{12}{\epsilon_V} \frac{y^6}{(1+y^6)^2} - 1\right] \Bigg\}\,. \end{aligned} \tag{7.2.33}$$

The number of e-folds between the initial time t_i and a given time t is given by

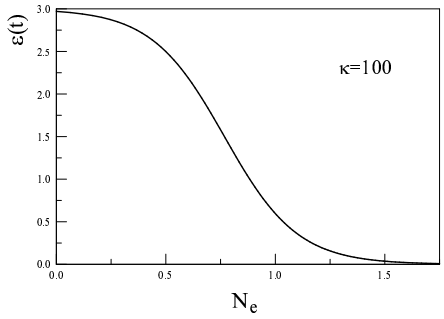
$$N_e(t;t_i) = \int_{t_i}^{t} H(t')\, dt' = \frac{1}{3}\, \ln\left[\sqrt{\kappa}\, \frac{(1-x^6(t))}{2x^3(t)}\right], \tag{7.2.34}$$

with a total number of e-folds between the beginning of the fast roll stage at $t = t_i$ and the onset of slow roll at t_{sr} given by

$$N_e(t_i;t_{sr}) = \frac{1}{6}\, \ln\left[\frac{3\kappa}{\epsilon_V}\right]. \tag{7.2.35}$$

Fig. (34) shows ε as a function of N_e for $\kappa = 100, \epsilon_V = 0.008$, inflation begins at $N_e \simeq 0.5 - 0.8$ and slow roll begins at $N_e \simeq 1.37 - 1.75$. We find that this is the typical behavior for $1 \leq \kappa \leq 100$, namely for a wide range of fast roll initial conditions, the inflationary stage begins fairly soon $N_{e,inf} \lesssim 1$ and the fast roll stage lasts $\lesssim 1.7$ e-folds.

The latest results from the Planck collaboration[63] confirm a $5 - 10\%$ suppression of power for $l \lesssim 40$ with $2.5 - 3\,\sigma$ significance. Recently in ref.[283] a detailed study of the



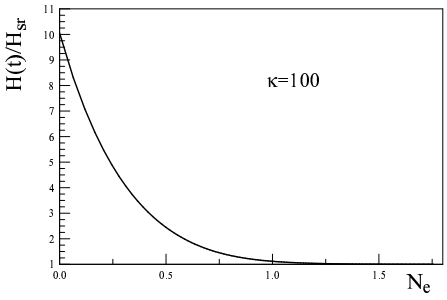


Figure 34: $\varepsilon(t)$ and $H(t)/H_{sr}$ as a function of the number of e-folds from the beginning of fast roll, for $\kappa = 100$ for $\epsilon_V = 0.008$. Inflation starts at $N_e \simeq 0.5$, slow roll starts at $N_e \lesssim 1.75$.

impact of the fast-roll stage on the suppression of the low multipoles has been reported. The results of this reference are independent of the inflaton potential and suggest that a $5-10\%$ suppression of the quadrupole is consistent with a fast roll stage with a ratio of kinetic to potential contributions $10 \lesssim \kappa \lesssim 100$. These results confirm more generally previous results based on particular realizations of the inflaton potential[273, 274, 275, 276, 277].

7.3 INITIAL CONDITIONS FROM A PRE-SLOW ROLL STAGE:

Our goal is to understand how infrared aspects of light scalar fields with mass $M \ll H$, are modified by the "fast-roll" stage, therefore in this and following sections we focus on "test" scalar fields, not necessarily the inflaton field.

The quantization of a generic minimally coupled massive scalar field is achieved by writing

$$\phi(\vec{x},\eta) = \frac{1}{C(\eta)}\frac{1}{\sqrt{V}}\sum_{\vec{k}}\left[\alpha_{\vec{k}}\, S(k,\eta)\, e^{i\vec{k}\cdot\vec{x}} + \alpha^{\dagger}_{\vec{k}}\, S^{*}(k,\eta)\, e^{-i\vec{k}\cdot\vec{x}}\right] , \tag{7.3.1}$$

where the operators $\alpha_{\vec{k}},\ \alpha^{\dagger}_{\vec{k}}$ obey the usual canonical commutation relations, and the mode functions $S_\phi(k,\eta)$ are solutions of

$$\left[\frac{d^2}{d\eta^2}+k^2-W(\eta)\right]S(k,\eta)=0\ \ ;\ \ W(\eta)=\frac{C''(\eta)}{C(\eta)}-M^2\,C^2(\eta)\,. \tag{7.3.2}$$

This is a Schrödinger equation, with η the coordinate, k^2 the energy and $W(\eta)$ a potential that depends on the coordinate η. The full dynamics of the inflaton field during the fast roll stage yields the potential

$$W(\eta)=\frac{C''}{C}-M^2\,C^2(\eta)=a[\ddot{a}+H\dot{a}]-M^2\,a^2(t)=2a^2H^2\Big[1-\frac{3}{2}\,\Delta-\frac{\varepsilon(t)}{2}\Big]\,, \tag{7.3.3}$$

where we have introduced

$$\Delta=\frac{M^2}{3H^2}\ll 1\,. \tag{7.3.4}$$

During slow roll inflation the potential $\varepsilon=\epsilon_V$ and

$$a^2(t)H^2(t)=\frac{1}{\eta^2}\,(1+2\epsilon_V) \tag{7.3.5}$$

therefore, during slow roll $W(\eta)$ becomes

$$W(\eta)=\frac{\nu^2-\frac{1}{4}}{\eta^2}\ , \tag{7.3.6}$$

where to leading order in slow roll parameters

$$\nu=\frac{3}{2}+\epsilon_V-\Delta\,. \tag{7.3.7}$$

Therefore during the full dynamics of the inflation including the fast roll stage we write

$$W(\eta)\equiv\mathcal{V}(\eta)+\frac{\nu^2-\frac{1}{4}}{\eta^2} \tag{7.3.8}$$

where the potential

$$\mathcal{V}(\eta)=W(\eta)-\frac{2}{\eta^2}\Big[1+\frac{3\epsilon_V}{2}-\frac{3\Delta}{2}\Big]\,. \tag{7.3.9}$$

The potential is calculated parametrically in terms of the variable x introduced in (7.2.26) and a, H, η all functions of x given by the expressions (7.2.28,7.2.29,7.2.33). Figure (35) shows the typical potentials for $\kappa=10,100; \epsilon_V=0.008; \Delta=0.01$. We studied the potentials for a wide range of values of ϵ_V, Δ and κ with qualitatively the same features.

The potentials are always *negative* and qualitatively of the same form with very small variations for fixed κ the (negative) amplitude of the potential increases with increasing κ. For both $\epsilon_V; \Delta \ll 1$ the potential is quite insensitive to their values and is mainly determined by the ratio κ.

These results are in general agreement with those of refs. [275, 276, 277] and more recently in ref.[283] an more detailed analysis confirmed the robustness of the main features of the pre-slow roll stage quite independently of the inflationary potential (provided the potential is smooth enough to be consistent with slow roll).

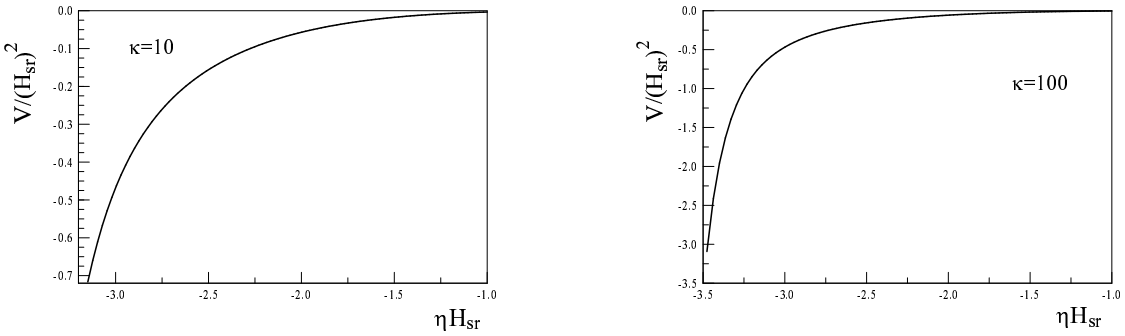


Figure 35: Potentials for $\kappa = 10; 100$; $\epsilon_V = 0.008$; $\Delta = 0.01$ as a function of η from the beginning of fast roll.

The solution of the mode equations with Bunch-Davies initial conditions for sub horizon modes obey the condition

$$S(k;\eta) \rightarrow \frac{e^{-ik\eta}}{\sqrt{2k}} \quad ; \quad -k\eta \rightarrow \infty \,, \tag{7.3.10}$$

and up to an overall phase are given by

$$S(k;\eta) \equiv g_\nu(k,\eta) = \sqrt{\frac{-\pi\eta}{4}}\, H^{(1)}_\nu(-k\eta) \,, \tag{7.3.11}$$

these mode functions satisfy the Wronskian condition

$$\mathcal{W}[g,g^*] = g^{'}_\nu(k,\eta)\, g^*_\nu(k,\eta) - g^{*\,'}_\nu(k,\eta)\, g_\nu(k,\eta) = -i \,. \tag{7.3.12}$$

When field quantization is carried out with these mode functions the vacuum state $|0\rangle_{BD}$ annihilated by the operators $\alpha_{\vec{k}}$ is the Bunch-Davies vacuum. However, the most general solution in the slow-roll regime can be written as

$$S(k;\eta) = A_k g_\nu(k,\eta) + B_k g^*_\nu(k,\eta) \tag{7.3.13}$$

where $A_k; B_k$ are Bogoliubov coefficients. For the creation and annihilation operators to obey standard commutation relations it follows that these general combinations must obey the Wronskian condition

$$\mathcal{W}[S,S^*] = -i = \mathcal{W}[g,g^*]\Big[|A_k|^2 - |B_k|^2\Big] \tag{7.3.14}$$

from which it follows that the Bogoliubov coefficients must obey the constraint

$$|A_k|^2 - |B_k|^2 = 1\,. \tag{7.3.15}$$

The relation between quantization with the mode functions $S(k;\eta)$ with general initial conditions, and the more familiar Bunch-Davies case with the mode functions g_ν (7.3.11) is obtained from the expansion of the Fourier components of the relevant fields, namely the field can be expanded in either set with corresponding annihilation and creation operators, for example for a scalar field

$$\frac{1}{\sqrt{V}}\sum_k a_{\vec{k}} g_\nu(k,\eta) e^{i\vec{k}\cdot\vec{x}} + a^\dagger_{\vec{k}} g^*_\nu(k,\eta) e^{-i\vec{k}\cdot\vec{x}} = \frac{1}{\sqrt{V}}\sum_k \alpha_{\vec{k}} S(k,\eta) e^{i\vec{k}\cdot\vec{x}} + \alpha^\dagger_{\vec{k}} S^*(k,\eta) e^{-i\vec{k}\cdot\vec{x}} \tag{7.3.16}$$

where $a_{\vec{k}}|0\rangle_{BD} = 0$ defines the Bunch-Davies vacuum and $\alpha_{\vec{k}}|0\rangle_\alpha$ defines the vacuum with the general initial conditions. The relation between the creation and annihilation operators is obtained from the Wronskian conditions, it is given by

$$\alpha_{\vec{k}} = A^*_k a_{\vec{k}} - B^*_k a^\dagger_{-\vec{k}} \;\;;\;\; \alpha^\dagger_{\vec{k}} = A_k a^\dagger_k - B_k a_{-\vec{k}}\,. \tag{7.3.17}$$

The Bogoliubov coefficients have been discussed in the literature[241, 245, 242, 243, 244] and an interpretation can be furnished by considering the action of the α number operator on the Bunch-Davies vacuum. It is easily shown that

$$ {}_{BD}\langle 0|\alpha_k^\dagger \alpha_k|0\rangle_{BD} = |B_k|^2 \tag{7.3.18}$$

which suggests the interpretation that $|B_k|$ is the number of α-vacuum particles in the Bunch Davies vacuum.

The power spectra for scalar field fluctuations (ϕ),

$$\mathcal{P}(k) = \frac{k^3}{2\pi^2}\left|\frac{S(k;\eta)}{C(\eta)}\right|^2 \tag{7.3.19}$$

Evaluating these power spectra a few e-folds after horizon crossing $-k\eta \ll 1$ and using that in this regime $H_\nu^{(1)}(-k\eta) \simeq i\, Y_\nu(-k\eta)$ it follows that for $-k\eta \ll 1$ the general solution of the form (7.3.13) is given by

$$S(k;\eta) = i\sqrt{\frac{-\pi\eta}{4}}\, Y_\nu(-k\eta)\big[A_k - B_k\big]\ , \tag{7.3.20}$$

therefore the power spectra becomes

$$\mathcal{P}(k) = \mathcal{P}^{BD}(k)\,\mathcal{T}(k)\,, \tag{7.3.21}$$

where $\mathcal{P}^{BD}(k)$ are the power spectra for Bunch-Davies modes $g_\nu(k;\eta)$, namely for $A_k = 1; B_k = 0$, and

$$\mathcal{T}(k) = \big|A_k - B_k\big|^2 \tag{7.3.22}$$

is a transfer function that encodes the non-Bunch-Davies initial conditions for the respective perturbations.

The main question is precisely what is the origin of $\mathcal{T}(k)$ and what are the properties for small and large k.

In references [275, 276, 277] and more recently in ref.[283] the modifications of the mode equations during the fast-roll stage were invoked as a *possible* origin of the coefficients A_k, B_k

was considered to be the brief kinetic dominated "fast-roll" stage just prior to the near de Sitter slow roll stage discussed above.

Here we pursue this line of argument and consider this possibility in detail, in particular focusing on the superhorizon limit of the transfer function $\mathcal{T}(k)$ (7.3.22) for light "test" scalar fields, namely with $\Delta \ll 1$.

The full dynamical evolution of the inflaton leads to a modification of the mode equations (7.3.2) where $W(\eta)$ is now given by (7.3.9) in terms of the potential $\mathcal{V}(\eta)$. As shown in figure (35) this potential is localized in η in a narrow range prior to the slow roll phase[275, 276, 277, 283], namely in the mode equations (7.3.2) $W(\eta)$ is written as

$$W(\eta) = \mathcal{V}(\eta) + \frac{\nu^2 - 1/4}{\eta^2} \quad ; \quad \mathcal{V}(\eta) = \begin{cases} \neq 0 \text{ for } -\infty < \eta < \eta_{sr} \\ 0 \text{ for } \eta_{sr} < \eta \end{cases} . \tag{7.3.23}$$

where η_{sr} determines the beginning of the slow roll stage when $\epsilon_V \ , \ \eta_V \ll 1$ (see figure (35)).

Rather than studying the behavior of the Bogoliubov coefficients numerically for different values of the parameters, we now exploit the similarity with a quantum mechanical potential problem and implement methods from the quantum theory of scattering to obtain the *general* behavior on $\mathcal{T}(k)$ for small and large wavevectors based solely on the fact that the potential is *negative* and localized. These are *generic* features of the potentials $\mathcal{V}(\eta)$ as consequence of the brief fast roll stage prior to slow roll.

The mode equation (7.3.2) can now be written as

$$\Big[\frac{d^2}{d\eta^2} + k^2 - \frac{\nu^2 - 1/4}{\eta^2}\Big] S(k;\eta) = \mathcal{V}(\eta) S(k;\eta) \,, \tag{7.3.24}$$

which can be converted into an integral equation via the retarded Green's function $G_k(\eta, \eta')$ obeying

$$\left[\frac{d^2}{d\eta^2} + k^2 - \frac{\nu^2 - \frac{1}{4}}{\eta^2}\right] G_k(\eta,\eta') = \delta(\eta - \eta') \quad ; \quad G_k(\eta,\eta') = 0 \text{ for } \eta' > \eta \,. \tag{7.3.25}$$

This Green's function is given by

$$G_k(\eta,\eta') = i\left[g_\nu(k;\eta)\, g^*_\nu(k;\eta') - g_\nu(k;\eta')\, g^*_\nu(k;\eta)\right] \Theta(\eta - \eta') \quad , \tag{7.3.26}$$

where $g_\nu(k;\eta)$ is given by eq.(7.3.10). The solution of (7.3.24) with boundary conditions corresponding to Bunch-Davies modes deep inside the horizon obeys the Lippman-Schwinger integral equation familiar from scattering theory,

$$S(k;\eta) = g_\nu(k;\eta) + \int_{-\infty}^{0} G_k(\eta,\eta')\ \mathcal{V}(\eta')\ S(k;\eta')\ d\eta'\ . \tag{7.3.27}$$

With the Green's function given by (7.3.25) this solution can be written as

$$S(k;\eta) = A_k(\eta)\, g_\nu(k;\eta) + B_k(\eta)\, g^*_\nu(k;\eta)\,, \tag{7.3.28}$$

where

$$A_k(\eta) \;=\; 1 + i\int_{-\infty}^{\eta} \mathcal{V}(\eta')\, g^*_\nu(k;\eta')\, S(k;\eta')\, d\eta' \tag{7.3.29}$$

$$B_k(\eta) \;=\; -i\int_{-\infty}^{\eta} \mathcal{V}(\eta')\, g_\nu(k;\eta')\, S(k;\eta')\, d\eta'\,. \tag{7.3.30}$$

For a potential $\mathcal{V}(\eta)$ that is localized prior to the slow roll stage (see fig. 35), and for $\eta > \eta_{sr}$ we can safely replace the upper limit of the integrals $\eta \to 0$ and during the slow roll stage the solution (7.3.28) becomes

$$S(k;\eta) = A_k\, g_\nu(k;\eta) + B_k\, g^*_\nu(k;\eta) \quad ; \quad A_k \equiv A_k(\eta=0)\ ;\ B_k \equiv B_k(\eta=0)\,. \tag{7.3.31}$$

This expression clearly suggests that mode functions with general initial conditions follow from pre-slow-roll stage wherein the inflaton zero mode undergoes rapid dynamical evolution. Refs.[283] provides a thorough numerical study of the potential independently of the inflaton potential (see figs. in this reference).

We now pursue an analytic understanding of the transfer function $\mathcal{T}(k)$ both for super and subhorizon modes quite generically without specifying particular values of $\kappa; \epsilon_V; \Delta$ but based *solely* on the fact that the potential $\mathcal{V}(\eta)$ is localized and negative.

We first note that the η dependent Bogoliubov coefficients (7.3.29,7.3.30) satisfy the relation

$$g_\nu(k;\eta)A'_k(\eta) + g^*_\nu(k;\eta)B'_k(\eta) = 0\,, \tag{7.3.32}$$

which implies the following relation between Wronskians

$$\mathcal{W}[S,S^*] = \mathcal{W}[g_\nu, g^*_\nu]\left(|A_k(\eta)|^2 - |B_k(\eta)|^2\right). \tag{7.3.33}$$

valid at *all times* not only during slow roll.

Secondly, inserting the relation (7.3.28) into the equations (7.3.29,7.3.30) leads to the coupled Fredholm integral equations

$$A_k(\eta) = 1 + i\int_{-\infty}^{\eta}\Big\{C(k;\eta')A_k(\eta') + D(k;\eta')B_k(\eta')\Big\}\,d\eta' \tag{7.3.34}$$

$$B_k(\eta) = -i\int_{-\infty}^{\eta}\Big\{C(k;\eta')B_k(\eta') + D^*(k;\eta')A_k(\eta')\Big\}\,d\eta'\,, \tag{7.3.35}$$

where the coefficient functions

$$C(k;\eta) = |g_\nu(k;\eta)|^2\,\mathcal{V}(\eta) \quad ; \quad D(k;\eta) = \big(g_\nu^*(k;\eta)\big)^2\,\mathcal{V}(\eta)\,. \tag{7.3.36}$$

Upon taking derivatives with respect to conformal time we find the coupled differential equations

$$A_k'(\eta) = iC(k;\eta)A_k(\eta') + iD(k;\eta)B_k(\eta) \quad ; \quad A_k(k;-\infty) = 1 \tag{7.3.37}$$

$$B_k'(\eta) = -iC(k;\eta)B_k(\eta) - iD^*(k;\eta)A_k(\eta) \quad ; \quad B_k(k;-\infty) = 0\,. \tag{7.3.38}$$

It is straightforward to confirm that these equations lead to the result

$$\frac{d}{d\eta}\Big(|A_k(\eta)|^2 - |B_k(\eta)|^2\Big) = 0\,, \tag{7.3.39}$$

which combined with the initial conditions in eqns. (7.3.37,7.3.38) yield the η-independent result

$$|A_k(\eta)|^2 - |B_k(\eta)|^2 = 1\,. \tag{7.3.40}$$

Along with the relation (7.3.33) this result implies that $\mathcal{W}[S,S^*] = -i$, namely the fields quantized with the Bunch-Davies modes and the modes $S(k;\eta)$ which are determined by the pre-slow roll stage are related by a canonical transformation.

Writing the coefficients $C(k;\eta); D(k;\eta)$ explicitly in terms of Bessel functions, it follows that

$$C(k;\eta) + D^*(k;\eta) = \left(\frac{-\pi\eta}{2}\right)\mathcal{V}(\eta)\Big[J_\nu^2(-k\eta) + iJ_\nu(-k\eta)Y_\nu(-k\eta)\Big]\,. \tag{7.3.41}$$

$$C(k;\eta) - D^*(k;\eta) = \left(\frac{-\pi\eta}{2}\right)\mathcal{V}(\eta)\Big[Y_\nu^2(-k\eta) - iJ_\nu(-k\eta)Y_\nu(-k\eta)\Big]\,. \tag{7.3.42}$$

The coupled set of linear differential equations (7.3.37,7.3.38) is difficult to solve analytically in general although the system is amenable to a straightforward numerical integration. However analytical progress can be made in two limits: a) the superhorizon limit $-k\eta \to 0$, b) subhorizon modes $-k\eta \gg 1$.

Superhorizon modes: For modes that crossed the horizon prior to the onset of the slow-roll phase and either during or prior to the stage where the inflaton field is evolving rapidly

$$J_\nu(-k\eta) \simeq \frac{\big(-k\eta/2\big)^\nu}{\nu\,\Gamma(\nu)} \quad ; \quad Y_\nu(-k\eta) \simeq -\frac{\Gamma(\nu)}{\pi}\,\big(-k\eta/2\big)^{-\nu}\,. \tag{7.3.43}$$

It proves convenient to define the combinations

$$F_\pm(k;\eta) = A_k(\eta) \pm B_k(\eta)\,, \tag{7.3.44}$$

obeying the coupled equations

$$F_-'(k;\eta) - \gamma(\eta) F_-(k;\eta) = i\pi\nu\gamma(\eta) J_\nu^2(-k\eta) F_+(k;\eta) \tag{7.3.45}$$

$$F_+'(k;\eta) + \gamma(\eta) F_+(k;\eta) = i\pi\nu\gamma(\eta) Y_\nu^2(-k\eta) F_-(k;\eta)\,, \tag{7.3.46}$$

where we have introduced

$$\gamma(\eta) = \left(\frac{-\eta}{2\nu}\right)\mathcal{V}(\eta)\,. \tag{7.3.47}$$

The equations (7.3.45,7.3.46) can be simplified by writing

$$F_\pm(k;\eta) = h_\pm(\eta)\, f_\pm(k;\eta) \quad ; \quad h_\pm(\eta) = \exp\left\{\mp\int_{-\infty}^{\eta} d\eta'\gamma(\eta')\right\}, \tag{7.3.48}$$

and defining

$$\tilde{j}(k;\eta) \equiv \pi\nu J_\nu^2(-k\eta)\, h_+^2(\eta) \quad ; \quad \pi\nu Y_\nu^2(-k\eta)\, h_-^2(\eta) = \frac{1}{\tilde{j}(k;\eta)}\,, \tag{7.3.49}$$

where we have used the limiting form (7.3.43) for superhorizon modes. With these definitions one finds the following set of coupled equations for the real and imaginary parts,

$$\mathrm{Re}\, f_-'(k;\eta) \;=\; -\gamma(\eta)\,\tilde{j}(k;\eta)\;\mathrm{Im}\, f_+(k;\eta) \tag{7.3.50}$$

$$\mathrm{Re}\, f_+'(k;\eta) \;=\; -\frac{\gamma(\eta)}{\tilde{j}(k;\eta)}\,\mathrm{Im}\, f_-(k;\eta)\,, \tag{7.3.51}$$

$$\mathrm{Im}\, f'_-(k;\eta) = \gamma(\eta)\, \tilde{j}(k;\eta)\ \mathrm{Re}\, f_+(k;\eta) \tag{7.3.52}$$

$$\mathrm{Im}\, f'_+(k;\eta) = \frac{\gamma(\eta)}{\tilde{j}(k;\eta)}\, \mathrm{Re}\, f_-(k;\eta)\,, \tag{7.3.53}$$

with the initial conditions

$$\mathrm{Re} f_\pm(k;\eta \to -\infty) \to 1 \quad ; \quad \mathrm{Im} f_\pm(k;\eta \to -\infty) \to 0\,. \tag{7.3.54}$$

Given the potential $\mathcal{V}(\eta)$ this set of equations lends itself to a simple numerical integration. However we can pursue further analytical understanding by writing them into an equivalent set of integral equations as follows: formally integrating (7.3.51,7.3.53) with the initial condition (7.3.54) and introducing the result into the equations for (7.3.50,7.3.52), we integrate with the initial condition (7.3.54) and obtain

$$\mathrm{Re} f_-(k;\eta) = 1 - \int_{-\infty}^{\eta} d\eta' \gamma(\eta') \tilde{j}(k;\eta') \int_{-\infty}^{\eta'} d\eta'' \frac{\gamma(\eta'')}{\tilde{j}(k;\eta'')}\, \mathrm{Re} f_-(k;\eta'') d\eta'' \tag{7.3.55}$$

$$\mathrm{Im} f_-(k;\eta) = \int_{-\infty}^{\eta} d\eta' \gamma(\eta') \tilde{j}(k;\eta') - \int_{-\infty}^{\eta} d\eta' \gamma(\eta') \tilde{j}(k;\eta') \int_{-\infty}^{\eta'} d\eta'' \frac{\gamma(\eta'')}{\tilde{j}(k;\eta'')}\, \mathrm{Im} f_-(k;\eta'') d\eta''\,. \tag{7.3.56}$$

Inserting the solutions to these integral equations into equations (7.3.50,7.3.52) yield the solutions for $f_+(k;\eta)$.

We can glean several important features from the integral equations (7.3.55,7.3.56):

- $\mathrm{Re}f_-(k;\eta)$ has a smooth $k \to 0$ limit as the factors $k^{2\nu}$ cancel between the $\tilde{j}$ in the numerator and denominator in the integral equations. Using the small argument expansion of Bessel functions we find that in the long-wavelength limit

$$\mathrm{Re}f_-(k;0) \simeq \mathrm{Re}f_-(0;0) + \mathcal{O}(k^2) + \cdots . \tag{7.3.57}$$

where $\mathrm{Re}f_-(0;0)$ is finite.

Since $\tilde{j}(k;\eta) \propto k^{2\nu}$ one notes that rescaling $\mathrm{Im}f_-(k;\eta) \equiv k^{2\nu}\,\mathrm{Im}\tilde{f}_-(k;\eta)$, it follows from eqn. (7.3.56) that $\mathrm{Im}\tilde{f}_-(k;\eta)$ has a finite limit as $k \to 0$ therefore we find that in the long wavelength limit

$$\mathrm{Im}f_-(k;\eta) \simeq \mathcal{C}\,k^{2\nu}\Big[1 + \mathcal{O}(k^2) + \cdots\Big] . \tag{7.3.58}$$

where $\mathcal{C}$ is a finite constant, therefore $\mathrm{Im}[A_{k=0}(\eta) - B_{k=0}(\eta)] = 0$. From the result $|A_k(\eta)|^2 - |B_k(\eta)|^2 = 1$ this implies that the real part $\mathrm{Re}[A_{k=0} - B_{k=0}]$ can *never vanish*. Because of the initial condition this combination begins positive ($= 1$) in the early past and *always remains positive* and the double integral in (7.3.55) is manifestly *positive* and finite leading to the conclusion that

$$\mathrm{Re}f_-(0;0) < 1 \quad ; \quad \mathrm{Im}f_-(0;0) = 0 . \tag{7.3.59}$$

From the result (7.3.58) above, and inserting this result in eqn. (7.3.51) we find that $\mathrm{Re}f_+(k;\eta)$ features a smooth long-wavelength limit with $\mathrm{Re}f_+(0;0)$ a finite constant. Inserting the result that $\mathrm{Re}f_-(k;\eta)$ is a regular function approaching a constant in the long-wavelength limit it follows that $\mathrm{Im}f_+(k;\eta) \propto k^{-2\nu}$ and features an infrared divergence in the long-wavelength limit. These results for $f_+(k;\eta)$ imply that in the long wavelength limit the *sum*

$$A_k + B_k \propto i\,k^{-2\nu} . \tag{7.3.60}$$

It is important to recognize how, in view of this result, the identity $|A_k(\eta)|^2 - |B_k(\eta)|^2 = 1$ is fulfilled in the long wavelength limit: from the results $\mathrm{Im}f_-(0;\eta) = 0$ and the long wavelengh limit $\mathrm{Im}f_+(k;\eta) \propto k^{-2\nu}$ it follows that in this limit $[\mathrm{Im}A_k(\eta)]^2 = [\mathrm{Im}B_k(\eta)]^2 \propto k^{-4\nu}$ and $[\mathrm{Re}A_k(\eta)]^2 \simeq \mathcal{O}(1)$; $[\mathrm{Re}B_k(\eta)]^2 \simeq \mathcal{O}(1)$ from which it follows that $|A_k(\eta)|^2 - |B_k(\eta)|^2 \simeq \mathcal{O}(1)$, namely the singular long wavelength behavior in the imaginary parts

of the Bogoliubov coefficients cancel out in the long-wavelength limit, leaving only the regular contributions in this limit.

During the slow-roll, near de Sitter stage the mode functions become

$$S(k;\eta) = \frac{1}{2}\sqrt{-\pi\,\eta}\left[(A_k + B_k)\,J_\nu(-k\eta) + i(A_k - B_k)\,Y_\nu(-k\eta)\right] \tag{7.3.61}$$

in the long-wavelength and long time limit, with the result that $A_k + B_k \propto ik^{-2\nu}$ and $A_k - B_k \simeq \mathcal{O}(1)$, it follows that

$$S(k;\eta) \simeq \left[a\,k^{-\nu}(-\eta)^{\frac{1}{2}+\nu} + b\,k^{-\nu}\left(A_{k=0} - B_{k=0}\right)(-\eta)^{\frac{1}{2}-\nu}\right], \tag{7.3.62}$$

where a, b are coefficients of $\mathcal{O}(1)$. Hence, although both terms are of the same order $\propto k^{-\nu}$ in the long wavelength limit, it is the second term that dominates well after horizon crossing and the power spectrum is determined by this term as anticipated above, see the discussion leading up to eqns. (7.3.21,7.3.22). In summary for long-wavelength modes at long time $\eta \to 0$ the mode functions can be approximated as

$$S(k;\eta) \simeq \frac{-i\,\Gamma(\nu)}{2\pi}(A_k - B_k)\,\sqrt{-\pi\,\eta}\,\left(\frac{2}{-\eta}\right)^{\nu} k^{-\nu}\,. \tag{7.3.63}$$

This result will be used in the analysis of infrared correlations in the next sections.

- The above results combined with equations (7.3.44) and (7.3.48) lead to

$$\mathrm{Re}[A_{k=0}(0) - B_{k=0}(0)] = \exp\left\{\int_{-\infty}^{0} d\eta'\gamma(\eta')\right\}\mathrm{Re}f_-(0;0) \quad ; \quad \mathrm{Im}[A_{k=0}(0) - B_{k=0}(0)] = 0\,, \tag{7.3.64}$$

hence,

$$\mathcal{T}(0) = \exp\left\{2\int_{-\infty}^{0} d\eta'\gamma(\eta')\right\}\left[\mathrm{Re}f_-(0;0)\right]^2. \tag{7.3.65}$$

Therefore for an *attractive potential* $\mathcal{V}(\eta) < 0$ it follows that $\gamma(\eta) < 0$ and

$$\mathcal{T}(0) < 1\,, \tag{7.3.66}$$

namely, for an *attractive potential* the long wavelength limit of the initial condition transfer function is smaller than 1 entailing a *suppression* of the power spectrum at long wavelengths. Since $\left[\mathrm{Re}f_-(0;0)\right]^2 \leq 1$ for the case of attractive potentials as found for

a fast-roll stage[275, 276, 277] an *upper bound* for the superhorizon limit of the initial condition transfer function is

$$\mathcal{T}(0) \leq \exp\left\{2\int_{-\infty}^{0} d\eta' \gamma(\eta')\right\}. \tag{7.3.67}$$

This analysis confirms more generally the numerical results obtained in refs.[275, 276, 277]. Furthermore using the small argument approximation of the Bessel functions with non-integer ν the integral equations (7.3.55,7.3.56) clearly show that

$$\mathcal{T}(k) \simeq \mathcal{T}(0) + \mathcal{O}(k^2) + \cdots \tag{7.3.68}$$

namely has a power series expansion in k at long wavelengths.

Subhorizon modes: For modes that remain inside the Hubble radius throughout inflation $-k\eta \gg 1$ the integral equation (7.3.27) can be consistently solved in a Born series. In the first Born approximation we replace $S(k;\eta) = g_\nu(k;\eta)$ in the integral equation (7.3.27) leading to

$$A_k(\eta) \simeq 1 + \frac{i}{2k}\int_{-\infty}^{\eta} \mathcal{V}(\eta')\, d\eta' \tag{7.3.69}$$

$$B_k(\eta) \simeq -\frac{e^{-i\pi\nu}}{2k}\int_{-\infty}^{\eta} e^{-2ik\eta'}\mathcal{V}(\eta')\, d\eta'\,. \tag{7.3.70}$$

where we have used that for subhorizon modes $g_\nu(k;\eta) \to e^{-i\frac{\pi}{2}(\nu+1/2)}/\sqrt{2k}$. The subhorizon limit of the coefficient B_k is strongly suppressed because the Fourier transform of the localized potential $\mathcal{V}$ falls of very fast as a function of k for large k as a consequence of the Riemann-Lebesgue lemma. An integration by parts dropping the surface terms because a) for large k the integrand at the lower limit averages out to zero and b) for $\eta > \eta_{sr}$ the integrand vanishes at the upper limit since $\mathcal{V}(\eta > \eta_{sr}) = 0$, yields that during the slow roll stage when $\mathcal{V}(\eta) = 0$

$$B_k(\eta) \simeq -i\frac{e^{-i\pi\nu}}{4k^2}\int_{-\infty}^{\eta} e^{-2ik\eta'}\mathcal{V}'(\eta')\, d\eta' \to |B_k(\eta)|^2 \lesssim \frac{1}{16k^4}\,. \tag{7.3.71}$$

This implies that for subhorizon modes

$$|A_k(0)|^2 - 1 = |B_k(0)|^2 \lesssim \frac{1}{k^4}\,, \tag{7.3.72}$$

therefore the number of Bunch-Davies particles falls off very fast at large (subhorizon) momenta and the general initial conditions do not affect the short distance and renormalization aspects. Therefore, for modes that are deep within the Hubble radius during most of the slow roll era, and therefore, where very deep inside the Hubble radius during the pre-slow roll era it follows that

$$\mathcal{T}(k) = 1 + \mathcal{O}(1/k^4) + \cdots . \tag{7.3.73}$$

Although the intermediate range of momenta must be studied numerically for definite realization of the pre-slow roll potentials there are several relevant consequences of the results obtained in the superhorizon and subhorizon limits:

- On the largest scales *today* corresponding to wavevectors that crossed the horizon ~ 60 e-folds before the end of inflation, the initial conditions set by a pre-slow roll rapid dynamical evolution of the inflaton yields a *suppression* of the power spectrum when the potential $\mathcal{V}(\eta)$ is attractive, this is the situation for a "fast-roll" stage as confirmed numerically in refs. [275, 276, 277]. This suppression *may* explain at least the large scale anomaly in the CMB reflected on the low power for the lowest multipoles[1].
- The effect of pre-slow roll initial conditions is negligible on small scales, those that crossed the horizon late or near the end of slow roll inflation. For example scales that reentered at the time of recombination imprinted on the first acoustic peaks, crossed out during $\simeq 10$ e-folds in the period lasting about 60 e-folds before the end of inflation. These modes were deep inside the Hubble radius during the pre-slow roll stage ($\gtrsim 60$ e-folds prior to the end of inflation) and their contribution to $\mathcal{T}(k)$ is strongly suppressed.

  This suggests that these initial conditions *may* suppress the power spectrum for the largest scales but do not modify the spectral index and do not introduce a significant running of the spectral index with wavevector.

Although this latter consequence must be studied in further detail numerically, we now focus on the impact of these type of initial conditions upon the infrared aspects of correlations for light scalar fields during de Sitter inflation, postponing a detailed analysis for

[1] Although it is unlikely to explain the low multipole alignment or large scale asymmetry.

curvature perturbations to further study. In particular, we have found that whereas individually the Bogoliubov coefficients feature large contributions for superhorizon momenta (as determined by the result for the sum $A_k + B_k \propto k^{-2\nu}$), the power spectrum is only sensitive to the *difference* and is smooth with a finite limit for superhorizon momenta, thus the question remains: are there any other infrared sensitive quantities that may feature a stronger dependence on initial conditions?. We study below the following infrared aspects: the self-consistent generation of mass and the decay width of single particle states during de Sitter inflation, both are consequences of a strong infrared enhancement of nearly massless fields in inflationary cosmology, and cross-correlation between sub and superhorizon modes in the decay products.

7.4 INFRARED ASPECTS OF SCALAR FIELD CORRELATIONS.

Our goal is to study the influence of initial conditions on infrared aspects of scalar field correlators, in particular to assess how initial conditions arising from the pre-slow roll stage modify the self-consistent mass generated by infrared fluctuations and also how they affect the decay of single particle states and cross-horizon correlations.

For the purposes of this work, only minimally coupled scalar field theories in a spatially flat de Sitter cosmology will be considered. The action for this field is given by

$$I = \int d^3x\, dt\, a^3(t) \left\{ \frac{1}{2}\dot{\phi}^2 - \frac{(\nabla\phi)^2}{2a^2} - V(\phi) \right\} \tag{7.4.1}$$

The potential under consideration will be of the form

$$V(\phi) = \frac{1}{2}M^2\phi^2 + \lambda\phi^p \quad ; \quad p = 3, 4 \tag{7.4.2}$$

Passing to conformal time and conformally rescaling the fields

$$a(t(\eta)) \equiv C(\eta) = \frac{-1}{H\eta} \quad ; \quad a(t)\phi(\vec{x}, t) \equiv \chi(\vec{x}, \eta)\,, \tag{7.4.3}$$

the action can be rewritten, after discarding surface terms, as

$$I = \int d^3x\, d\eta \left\{ \frac{1}{2}\left[\chi'^2 - (\nabla\chi)^2 - \mathcal{M}^2(\eta)\chi^2\right] - \lambda(C(\eta))^{4-p}\chi^p \right\} \tag{7.4.4}$$

$$\mathcal{M}^2(\eta) \equiv M^2C^2(\eta) - \frac{C''(\eta)}{C(\eta)} = \frac{1}{\eta^2}\left[\frac{M^2}{H^2} - 2\right] \tag{7.4.5}$$

where $' = d/d\eta$. The equations of motion for the Fourier modes in the non-interacting theory during the de Sitter stage become

$$\chi''_{\vec{k}}(\eta) + \left[k^2 - \frac{1}{\eta^2}\left(\nu^2 - \frac{1}{4}\right)\right]\chi_{\vec{k}}(\eta) = 0 \quad ; \quad \nu^2 = \frac{9}{4} - \frac{M^2}{H^2} \tag{7.4.6}$$

Furthermore, we focus on light, nearly massless fields with $M^2/H^2 \ll 1$ in exact de Sitter space time in which case it follows that $\epsilon_V = \eta_V = 0$ and

$$\nu = \frac{3}{2} - \Delta \quad ; \quad \Delta = \frac{M^2}{3H^2} + \cdots \ll 1\,. \tag{7.4.7}$$

Infrared divergences arising from the nearly masslessness of the fields are manifest as poles in Δ in the various correlation functions[199, 200, 296, 298, 175, 291], we will focus on the leading order infrared contributions arising from the poles in Δ.

In order to study the effect of initial conditions set by a pre-de Sitter stage, we now quantize the scalar field with the general mode functions (7.3.31),

$$\chi(\eta, x) = \frac{1}{\sqrt{V}}\sum_k \alpha_k S_\nu(k,\eta)e^{i\vec{k}\cdot\vec{x}} + \alpha_k^\dagger S_\nu^*(k,\eta)e^{-i\vec{k}\cdot\vec{x}} \tag{7.4.8}$$

where $S_1 = A_k g_\nu(k,\eta) + B_k g_\nu^*(k,\eta)$ and $\alpha|0_\alpha\rangle = 0$ defines the vacuum with general initial conditions and the Bunch-Davies mode functions are given by (7.3.11), and the coefficients A_k, B_k obey the relation (7.3.15).

Two results obtained in the previous section are relevant for the analysis that follows:

$$\mathcal{T}(k) = |A_k - B_k|^2 \xrightarrow{k \to 0} \mathcal{T}(0) + \mathcal{O}(k^2) + \cdots \tag{7.4.9}$$

$$|A_k| \xrightarrow{k \to \infty} 1 + \mathcal{O}(1/k^4) \quad ; \quad |B_k| \xrightarrow{k \to \infty} \mathcal{O}(1/k^2) \tag{7.4.10}$$

With $\mathcal{T}(k)$ a smooth function of k and $\mathcal{T}(0)$ given by (7.3.65).

7.4.1 Interaction Picture

The time evolution of interacting fields is handled in a straightforward manner. In the Schrodinger picture, a quantum state $|\Psi(\eta)\rangle$ obeys

$$i\frac{d}{d\eta}|\Psi(\eta)\rangle = H(\eta)|\Psi(\eta)\rangle \tag{7.4.11}$$

where the Hamiltonian $H(\eta)$ is explicitly a function of η in an expanding cosmology. Defining the time evolution operator, this has the formal solution

$$i\frac{d}{d\eta}U(\eta,\eta_0) = H(\eta)U(\eta,\eta_0) \quad ; \quad U(\eta_0,\eta_0) = 1 \tag{7.4.12}$$

so that $|\Psi(\eta)\rangle = U(\eta,\eta_0)|\Psi(\eta_0)\rangle$. The Hamiltonian can be separated into free and interacting pieces, $H(\eta) = H_0(\eta) + H_i(\eta)$, where H_0 is the non-interaction Hamiltonian. Defining the time evolution operator for the free theory, $U_0(\eta,\eta_0)$, so that

$$i\frac{d}{d\eta}U_0(\eta,\eta_0) = H_0(\eta)U_0(\eta,\eta_0) \quad ; \quad i\frac{d}{d\eta}U_0^{-1}(\eta,\eta_0) = -U_0^{-1}(\eta,\eta_0)H_0(\eta) \quad ; \quad U(\eta_0,\eta_0) = 1 \tag{7.4.13}$$

From here, the interaction picture may be defined in the usual manner as

$$|\Psi(\eta)\rangle_I = U_I(\eta,\eta_0)|\Psi(\eta_0)\rangle_I = U_0^{-1}(\eta,\eta_0)|\Psi(\eta)\rangle \tag{7.4.14}$$

so that $U_I(\eta,\eta_0)$ is the interaction picture time evolution operator such that

$$\frac{d}{d\eta}U_I(\eta,\eta_0) = -iH_I(\eta)U_I(\eta,\eta_0) \quad ; \quad U_I(\eta_0,\eta_0) = 1 \quad ; \quad H_I(\eta) = U_0^{-1}(\eta,\eta_0)H_i(\eta)U_0(\eta,\eta_0) \tag{7.4.15}$$

For the interactions that will be considered here, the interaction Hamiltonian is given explicitly by

$$H_I(\eta) = \frac{\lambda}{(-H\eta)^{4-p}}\int d^3x(\chi(\vec{x},\eta))^p \tag{7.4.16}$$

To leading order in λ, the standard solution in perturbation theory is

$$U_I(\eta,\eta_0) = 1 - i\int_{\eta_0}^{\eta} d\eta' H_I(\eta') + ... \tag{7.4.17}$$

7.4.2 The infrared contribution to the tadpole:

The tadpole, $\langle 0|\chi^2(\vec{x},\eta)|0\rangle$ with $|0\rangle$ being the vacuum with non-Bunch Davies initial conditions, will play an important role in the following discussion. It is given by

$$\langle 0|\,\chi^2(\vec{x},\eta)|\,0\rangle = \int \frac{d^3k}{(2\pi)^3}|\,S(k,\eta)|^2\,. \tag{7.4.18}$$

Our goal is to extract the most relevant infrared contributions. In order to understand the influence of the Bogoliubov coefficients $A_k; B_k$ determined by the initial conditions, we revisit the evaluation of the tadpole for the Bunch-Davies case, namely $A_k = 1; B_k = 0$, $S(k;\eta) = g_\nu(k;\eta)$ to highlight the origin of the most infrared relevant contributions. In this case making a change of variables $y = -k\eta$ the tadpole is given by

$${}_{BD}\langle 0|\,\chi^2(\vec{x},\eta)|\,0\rangle_{BD} = \frac{1}{8\pi\,\eta^2}\int_0^{\Lambda_p/H}\frac{dy}{y}\,y^3|H_\nu^{(1)}(y)|^2 \tag{7.4.19}$$

where we have introduced an ultraviolet cutoff in physical coordinates. To isolate the infrared divergences for $\Delta \ll 1$ we write the integral above as

$$\int_0^{\Lambda_p/H}\frac{dy}{y}\,y^3|H_\nu^{(1)}(y)|^2 = \int_0^{\mu_p/H}\frac{dy}{y}\,y^3|H_\nu^{(1)}(y)|^2 + \int_{\mu_p/H}^{\Lambda_p/H}\frac{dy}{y}\,y^3|H_\nu^{(1)}(y)|^2 \tag{7.4.20}$$

with $\mu_p \to 0$ an infrared physical cutoff. For the first integral we use $\nu = 3/2 - \Delta$ with $0 < \Delta \ll 1$ and

$$z^3\,\left|H_\nu^{(1)}(z)\right|^2 \overset{z\to 0}{=} \left[\frac{2^\nu\,\Gamma(\nu)}{\pi}\right]^2\,z^{2\,\Delta} \tag{7.4.21}$$

thus $\Delta > 0$ regulates the infrared behavior of the tadpole and the first integral yields

$$\int_0^{\frac{\mu_p}{H}}\frac{dz}{z}\,z^3\,\left|H_\nu^{(1)}(z)\right|^2 = \frac{2}{\pi}\left[\frac{1}{2\,\Delta} + \frac{\mu_p^2}{2H^2} + \gamma - 2 + \ln\frac{2\mu_p}{H} + \mathcal{O}(\Delta)\right]\,, \tag{7.4.22}$$

where we have displayed the pole in Δ and the leading infrared logarithm. In the second integral in (7.4.20) we set $\nu = 3/2$ and combining its result with (7.4.22) we find that the

dependence on the infrared cutoff μ_p cancels in the limit $\mu_p \to 0$ leading to the following final result for the tadpole with Bunch-Davies vacuum

$$ {}_{BD}\langle 0|\,\chi^2(\vec{x},\eta)\,|0\rangle_{BD} = \frac{1}{8\pi^2\,\eta^2}\left[\frac{{\Lambda_p}^2}{H^2} + 2\ln\frac{\Lambda_p}{H} + \frac{1}{\Delta} + 2\,\gamma - 4 + \mathcal{O}(\Delta)\right] , \tag{7.4.23} $$

While the quadratic and logarithmic *ultraviolet* divergences are regularization scheme dependent, the pole in Δ arises from the infrared behavior and is independent of the regularization scheme. In particular this pole coincides with that found in the expression for $<\phi^2(\vec{x},t)>$ in refs.[175, 199, 200, 202]. The *ultraviolet divergences*, in whichever renormalization scheme, require that the effective field theory be defined to contain *renormalization counterterms* in the bare effective Lagrangian, for the tadpole this counterterm is of the form $\chi(\eta)\,J(\eta)$ and $J(n)$ is required to cancel the ultraviolet divergences. Thus, the *renormalized* tadpole in the Bunch-Davies vacuum is given by

$$ \mathcal{I}_{BD}(\eta) \equiv {}_{BD}\langle 0|\,\chi^2(\vec{x},\eta)\,|0\rangle^{ren}_{BD} = \frac{1}{8\pi^2\,\eta^2}\,\frac{1}{\Delta}\,\left[1+\cdots\right] , \tag{7.4.24} $$

where the dots stand for higher order terms in $\Delta \ll 1$.

We are now in position to understand the effect of non-Bunch-Davies initial conditions. The most infrared divergent contribution is determined by superhorizon modes for which $g_\nu(k;\eta) \simeq i\sqrt{-\pi\eta}\,Y_\nu(-k\eta)/2$ hence

$$ |S(k;\eta)|^2 \simeq \frac{-\pi\eta}{4} Y^2_\nu(-k\eta)\,\mathcal{T}(k) \quad ; \quad -k\eta \ll 1 \tag{7.4.25} $$

the fast fall off of the Bogoliubov coefficients with large k entails that the ultraviolet behavior of the tadpole is the same as in Bunch-Davies vacuum so that renormalization of the tadpole proceeds just as in the Bunch-Davies case. The pole in Δ in (7.4.23) arises from a narrow band of superhorizon wavevectors with the infrared cutoff $\mu \to 0$. The results of the previous section show that for superhorizon wavevectors $\mathcal{T}(k) = \mathcal{T}(0) + \mathcal{O}(k^2) + \cdots$ is a smooth function of k with $\mathcal{T}(0)$ given by (7.3.65). Therefore, to obtain the leading order infrared contribution for $\Delta \ll 1$ we replace $\mathcal{T}(k) \to \mathcal{T}(0)$ in (7.4.25) because the higher powers of k in $\mathcal{T}(k)$ yield terms that are subleading for $\Delta \ll 1$. Furthermore since for large k we found

that $\mathcal{T}(k) \lesssim 1/k^4$ the ultraviolet divergences of the tadpole are the same as for the Bunch-Davies case and renormalization is achieved in the same manner as with Bunch-Davies initial conditions.

Therefore for general initial conditions set during a pre-slow roll stage we obtain

$$\mathcal{I}(\eta) \equiv \langle 0|\, \chi^2(\vec{x},\eta)|\, 0\rangle^{ren} = \frac{1}{8\pi^2\,\eta^2}\,\frac{\mathcal{T}(0)}{\Delta}\,\left[1+\cdots\right], \tag{7.4.26}$$

Although this discussion was focused on the tadpole, similar arguments will allow to extract the leading infrared contributions in other correlators, the main point is that the leading infrared divergences that are responsible for poles in $\Delta \ll 1$ arise from a small band of superhorizon wavevectors for which $\mathcal{T}(k) \simeq \mathcal{T}(0)$.

7.4.3 Self Consistent Mass Generation

7.4.3.1 ϕ^3 **theory** For this particular field theory, radiative corrections will induce a non zero expectation value of the field in the "dressed" vacuum. At leading order for a general interaction Hamiltonian, the dressed vacuum evolves in time as

$$|\tilde{0}(\eta)\rangle = U(\eta,\eta_o)|\tilde{0}(\eta_o)\rangle \simeq \left(1 - i\int_{\eta_o}^{\eta} d\eta^{'} H_I(\eta^{'}) + ...\right)|\tilde{0}(\eta_o)\rangle \tag{7.4.27}$$

so that, to leading order, the expectation value of the field is given by

$$\begin{aligned}\langle\tilde{0}(\eta)|\,\chi(y,\eta)|\,\tilde{0}(\eta)\rangle &- \langle\tilde{0}(\eta_o)|\,\chi(y,\eta)|\,\tilde{0}(\eta_o)\rangle \equiv \delta\langle\chi(y,\eta)\rangle \\ &= i\langle\tilde{0}(\eta_o)|\int_{\eta_o}^{\eta} d\eta^{'}\,[H_I(\eta'),\chi(y,\eta)]\,|\tilde{0}(\eta_o)\rangle\end{aligned} \tag{7.4.28}$$

Specializing to $g\phi^3$ theory results in

$$\delta\langle\chi\rangle = 3i\lambda\int_{\eta_o}^{\eta} d\eta\,' C(\eta')\int d^3x[\chi(x,\eta\,'),\chi(y,\eta)]\langle\tilde{0}(\eta_o)|\chi^2(x,\eta')|\tilde{0}(\eta_o)\rangle \tag{7.4.29}$$

where the commutator is readily evaluated using the expansion of the field and creation / annhilation operator commutation relations, the result being

$$\int d^3x[\chi(x,\eta'),\chi(y,\eta)] = \left[S(k;\eta')S^*(k;\eta) - S^*(k;\eta')S(k;\eta)\right]_{k=0} \tag{7.4.30}$$

This is readily evaluated using the limiting form of Bessel functions and it can be shown that

$$\begin{aligned}
&S(k,\eta')S^*(k,\eta) = \\
&\frac{-\pi\sqrt{\eta\eta'}}{4}\Bigg\{(|A_k|^2+|B_k|^2)\left(\frac{1}{\Gamma^2(\nu+1)}\left(\frac{k\eta\eta'}{4}\right)^{\nu}+\frac{\Gamma^2(\nu)}{\pi^2}\left(\frac{k\eta\eta'}{4}\right)^{-\nu}\right)+ \\
&(A_kB_k^*+B_kA_k^*)\left(\frac{1}{\Gamma^2(\nu+1)}\left(\frac{k^2\eta\eta'}{4}\right)-\frac{\Gamma^2(\nu)}{\pi^2}\left(\frac{k\eta\eta'}{4}\right)^{-\nu}\right)+ \\
&(A_kB_k^*-B_kA_k^*)\left(\frac{-i}{\pi\nu}\right)\left(\left(\frac{\eta}{\eta'}^{-\nu}\right)+\left(\frac{\eta'}{\eta}^{-\nu}\right)\right) \\
&(|A_k|^2-|B_k|^2)\left(\frac{-i}{\pi\nu}\right)\left[\left(\frac{\eta'}{\eta}\right)^{-\nu}-\left(\frac{\eta}{\eta'}\right)^{-\nu}\right]\Bigg\},
\end{aligned} \tag{7.4.31}$$

note that the first three terms would diverge in the long wavelength limit, however these are all *real*, and $S(k,\eta')S^*(k,\eta) - S^*(k,\eta')S(k,\eta) = 2i\,\mathrm{Im}(S(k,\eta')S^*(k,\eta))$, hence these terms cancel in the expectation value. Since $|A_k|^2 - |B_k|^2 = 1$, the commutator becomes

$$\int d^3x[\chi(x,\eta'),\chi(y,\eta)] = \frac{i}{2\nu}(\eta^{\beta_+}\eta'^{\beta_-} - \eta^{\beta_-}\eta'^{\beta_+}) \quad ; \quad \beta_\pm = \frac{1}{2}\pm\nu \tag{7.4.32}$$

which is independent of the vacuum state.

Therefore, the full expression for the expectation value becomes

$$\delta\langle\chi\rangle = \frac{-3\lambda}{2\nu H}\int_{\eta_o}^{\eta}\frac{d\eta'}{\eta'}\left[\eta^{\beta_+}\eta'^{\beta_-} - \eta^{\beta_-}\eta'^{\beta_+}\right]\langle 0|\,\chi^2(\vec{x},\eta')|\,0\rangle \tag{7.4.33}$$

To leading order in Δ the renormalized tadpole contribution is given by (7.4.26)

$$\begin{aligned}\delta\langle\chi\rangle &= \frac{-3\lambda\mathcal{T}(0)}{16\pi\nu H\Delta}\int_{\eta_o}^{\eta}\frac{d\eta'}{\eta'^3}\left[\eta^{\beta_+}\eta'^{\beta_-}-\eta^{\beta_-}\eta'^{\beta_+}\right] \\ &= \frac{-\lambda\mathcal{T}(0)}{8\pi\Delta H\eta}\left(\frac{1}{\Delta}\left(1-\left(\frac{\eta}{\eta_o}\right)^{\Delta}\right)-\frac{1}{3}\left(1-\frac{\eta^3}{\eta_o^3}\right)\right)\end{aligned} \tag{7.4.34}$$

therefore to leading order in Δ and as $\eta/\eta_0 \to 0$ we find

$$\delta\langle\chi\rangle = \frac{-\lambda\mathcal{T}(0)}{8\pi H\Delta^2\eta} + \mathcal{O}(\Delta) \tag{7.4.35}$$

If the field initially has vanishing expectation value the interactions lead to a non-vanishing expectation value in the interacting ground state asymptotically given by

$$\langle 0|\chi(y,\eta)|0\rangle = \bar{\chi}(\eta) \to \frac{-\lambda}{8\pi^2 H\eta}\frac{\mathcal{T}(0)}{\Delta^2} + \mathcal{O}(\Delta) \tag{7.4.36}$$

Then the *unscaled* field obtains a constant expectation value for $\eta/\eta_o \to 0$,

$$\langle\phi(y,\eta)\rangle = \frac{1}{a(\eta)}\langle\phi(y,\eta)\rangle = \frac{\lambda}{8\pi^2}\frac{\mathcal{T}(0)}{\Delta^2} + \mathcal{O}(\Delta)\,. \tag{7.4.37}$$

This result which includes the effect of initial conditions is a generalization of that found in ref.[175] and is noteworthy because infrared effects lead to an asymptotic expectation value which is time independent, signaling the emergence of a non-trivial minimum of an effective action.

The emergence of a non-trivial expectation value and a mininum of the effective action implies that it is necessary to redefine the field shifting by this expectation value, namely

$$\chi(x,\eta) = \Psi(x,\eta) + \bar{\chi}(\eta) \quad ; \quad \langle\tilde{0}|\Psi(x,\eta)|\tilde{0}\rangle = 0 \tag{7.4.38}$$

This is the origin of the mechanism of self-consistent mass generation, for consider that the bare Lagrangian describes a massless scalar field with cubic interaction, shifting by the vacuum expectation value, the cubic term now written in terms of Ψ becomes

$$H_I = \int d^3x\left[\frac{1}{\eta^2}\frac{M^2}{2H^2}\Psi^2 - \frac{\lambda}{H\eta}\Psi^3\right] \tag{7.4.39}$$

where

$$\frac{1}{\eta^2}\frac{M^2}{2H^2} = -3\frac{\lambda}{H\eta}\bar{\chi}(\eta)\,. \tag{7.4.40}$$

This suggests a *self-consistent* mass generation mechanism by replacing $\bar{\chi}$ by the result (7.4.36), namely

$$\frac{1}{\eta^2}\frac{M^2}{2H^2} = \frac{3\lambda^2}{8\pi^2 H^2\eta^2}\frac{\mathcal{T}(0)}{\Delta^2}\left[1+\mathcal{O}(\Delta)\right] \tag{7.4.41}$$

since $\Delta = M^2/3H^2$ this is a self consistent condition with the result

$$M = H\sqrt{3}\left(\frac{\lambda}{2\pi H}\right)^{1/3}\left[\mathcal{T}(0)\right]^{1/6} \equiv M_{BD}\left[\mathcal{T}(0)\right]^{1/6} \tag{7.4.42}$$

where M_{BD} is the self-consistent mass obtained with Bunch-Davies initial conditions[175].

This is a noteworthy result, the strong infrared behavior leads to a self-consistent mass generation which is *non-analytic* in the transfer function for initial conditions.

7.4.3.2 ϕ^4 **theory** For this theory, the Lagrangian density is now $\mathcal{L}_I = -\lambda\chi^4$ and, as discussed previously, the expectation value of the field remains zero. As discussed in ref.[175] the mechanism of self consistent mass generation for a massless field is accomplished by introducing a mass term in the free Lagrangian and then subtracting it out again as a counterterm in the interaction part

$$\mathcal{L}_I = \frac{1}{2}C^2(\eta)M^2\chi^2 - \lambda\chi^4 \tag{7.4.43}$$

and requesting that the tadpole cancels the mass counterterm leading to a self-consistent condition akin to the Hartree resummation[296, 200, 297, 298], namely

$$\frac{M^2}{2H^2\eta^2} = 6\lambda\,\langle 0|\chi^2(x,\eta)|0\rangle \tag{7.4.44}$$

where the renormalized tadpole is given by (7.4.26), therefore to leading order in Δ, the self consistent mass becomes

$$M = H\left[\frac{9\lambda\,\mathcal{T}(0)}{2\pi^2}\right]^{1/4} \equiv M_{BD}\,\left[\mathcal{T}(0)\right]^{1/4}\,. \tag{7.4.45}$$

Again the Bunch-Davies case corresponds to $\mathcal{T}(0) = 1$ thus the self-consistent condition leading to dynamical mass generation from infrared divergences yields a non-analytic dependence of the generated mass upon the initial conditions.

The comparison between the infrared generated mass for Bunch-Davies initial conditions and the puzzling discrepancy obtained with other approaches[294, 290, 199, 200, 296, 298] has been discussed in ref.[175] (see the first reference).

7.4.4 Initial condition dependent anomalous dimensions:

The self-consistent mass generation through infrared divergences lead to the following expressions for Δ from the self-consistent solutions for cubic (3) and quartic (4) interactions respectively,

$$\Delta_{(3)} = \left[\frac{\lambda\sqrt{\mathcal{T}(0)}}{2\pi\, H}\right]^{\frac{2}{3}}, \tag{7.4.46}$$

$$\Delta_{(4)} = \left[\frac{\lambda\,\mathcal{T}(0)}{2\pi^2}\right]^{\frac{1}{2}}. \tag{7.4.47}$$

This result, in turn, implies that the power spectrum acquires non-perturbative initial condition-dependent *anomalous dimensions*, namely

$$\mathcal{P} \propto k^3\langle 0|\chi_{\vec{k}}(\eta)\,\chi_{-\vec{k}}(\eta)|0\rangle \propto k^{2\Delta}\,. \tag{7.4.48}$$

where Δ is given by (7.4.46,7.4.47) for cubic and quartic self-interactions respectively.

We highlight that for initial conditions determined by a fast-roll stage prior to slow roll, the long-wavelength power spectrum is suppressed and all the corrections from the initial conditions on self-consistent masses and anomalous dimensions are suppressed with respect to the Bunch-Davis result. Hence, initial conditions that *could* explain the anomalously low quadrupole in the CMB lead consistently to a suppression of all infrared effects, including the non-perturbatively generated masses and anomalous dimensions.

7.5 PARTICLE DECAY: WIDTH DEPENDENCE ON INITIAL CONDITIONS.

In an expanding cosmology, the lack of a global time-like Killing vector implies the lack of thresholds for particle decay (a consequence of energy-momentum conservation). Therefore, a single particle state of a field can decay into multiple particle states of the *same field* as discussed in refs.[190, 229] confirmed for heavy fields in ref.[191, 201] and more generally (and thoroughly) for a scalar theory with cubic interactions in[194]. The usual method to extract a decay rate in Minkowski space-time relies on energy-momentum conservation that leads to a transition probability that grows linearly in time at long times, namely a time-independent *decay rate*. The lack of energy conservation in an expanding cosmology prevents the usual implementation of what is, essentially, Fermi's Golden rule, instead the transition probability and ultimately the full time evolution of quantum states must be studied in detail.

In ref.[175] a non-perturbative field theoretical generalization of the Wigner-Weisskopf method to study the decay of single particle states was adapted to inflationary cosmology, and in ref.[175] this method was generalized and extended to obtain in a consistent manner both the infrared induced self-consistent masses and the time dependent decay width of particle states. The details of these methods have been explained in detail in refs.[175, 299, 198] and the reader is referred to these references for details. For self-consistency we give a brief summary of the method in appendix (7.8.2).

7.5.1 Transition Amplitude and Probability

To identify the corrections to masses and the decay widths, consider the interaction of a scalar fields through a cubic vertex. The interaction Hamiltonian is given by

$$H_i = \lambda \int d^3x\, a(t)^3 \phi^3 = \lambda C(\eta) \int d^3x\, \chi^3(x,\eta) \tag{7.5.1}$$

where the conformally rescaled fields have been used. Using the expansion of the field, Eq (7.4.8), the matrix element for process $\chi \to 2\chi$ can be readily obtained, it is given by

$$\mathcal{A}_{\chi\to\chi\chi} = \frac{-6i\lambda}{V^{1/2}} \int_{\eta_o}^{\eta} d\eta' C(\eta')\, S(k,\eta') S^*(k-q,\eta') S^*(q,\eta')\,, \tag{7.5.2}$$

and the total transition probability is given by

$$P_{\chi\to\chi\chi} = \sum_q |\mathcal{A}|^2 \equiv \int_{\eta_o}^{\eta} d\eta_1\, d\eta_2\, \Sigma(k,\eta_1,\eta_2) \tag{7.5.3}$$

where

$$\Sigma(k,\eta_1,\eta_2) = \frac{36\lambda^2}{H^2\eta_1\eta_2} \int \frac{d^3q}{(2\pi)^3}\, S^*(k,\eta_1) S(k,\eta_2) S(k-q,\eta_1) S^*(k-q,\eta_2) S(q,\eta_1) S^*(q,\eta_2)\,, \tag{7.5.4}$$

with the property

$$\Sigma(k,\eta_1,\eta_2) = \Sigma^*(k,\eta_2,\eta_1) \tag{7.5.5}$$

Inserting a factor of $1 = \theta(\eta_2-\eta_1) + \theta(\eta_1-\eta_2)$ in the integral and making use of (7.5.5) yields

$$P_{\chi\to\chi\chi}(k,\eta) = 2\int_{\eta_0}^{\eta} d\eta_2 \int_{\eta_0}^{\eta_2} d\eta_1 \mathrm{Re}\,[\Sigma(k,\eta_1,\eta_2)] \tag{7.5.6}$$

so the *transition rate* is easily identified to be

$$\Gamma(\eta) = \frac{d}{d\eta} P_{\chi\to\chi\chi}(k,\eta) = 2\int_{\eta_0}^{\eta} d\eta'\, \mathrm{Re}\,[\Sigma(k,\eta,\eta')]\ . \tag{7.5.7}$$

In Minkowski space time where energy-momentum conservation holds, the transition probability for a decaying state grows linearly (secularly) in time leading to a constant transition *rate* and an overall energy momentum delta function in the phase space integrals determining the kinematic reaction thresholds. Only when the transition probability grows with time is the process associated with the decay of the parent particle.

In an expanding cosmology there lack of energy conservation (energy momentum is covariantly conserved) leads to the lack of kinematic thresholds and the decay process $\chi \to 2\chi$ is allowed[229, 175]. In ref.[175] it is shown in detail non-perturbatively that an initial single particle state decays as

$$|\Psi(\eta)\rangle \propto |\Psi(\eta_0)\rangle \; e^{-\frac{1}{2}\int_{\eta_0}^{\eta} \Gamma(\eta')\, d\eta'} \,. \tag{7.5.8}$$

7.5.2 Cubic Vertex Decay Rate

In order to calculate the decay rate of $\chi \to 2\chi$ we need to evaluate $\Sigma(k,\eta_1,\eta_2)$ given by eqn. (7.5.4). We focus on the long time limit $\eta_1, \eta_2 \to 0$ and the leading order in Δ. The calculation is involved and has been carried out in detail for the case of Bunch-Davies initial conditions in ref.[175], the details of this calculation for general initial conditions with the Bogoliubov coefficients are relegated to appendix (7.8.1).

We find to leading order in Δ and in the long time limit,

$$\Sigma(k,\eta_1,\eta_2) = \frac{18\,\lambda^2\,\mathcal{T}(0)}{\pi^2\,H^2\,\Delta}\,\frac{|S(k,\eta_1)|^2}{(\eta_1)^2}\frac{|S(k,\eta_2)|^2}{(\eta_2)^2} + \mathcal{O}(\Delta^0) \tag{7.5.9}$$

The factor $\mathcal{T}(0)$ originates in the infrared region that yields the pole in Δ corresponding to one of the internal lines in the self energy, either $q \simeq 0$ or $q \simeq k$, within the band of superhorizon wavevectors. To leading order in Δ the self energy is purely real and the decay rate becomes

$$\Gamma(k;\eta) = \frac{36\lambda^2}{\pi^2 H^2}\frac{\mathcal{T}(0)}{\Delta}\frac{|S(k,\eta)|^2}{\eta^2}\int_{-\eta}^{-\eta_0} d(-\eta')\frac{|S(k,\eta')|^2}{(\eta')^2} \tag{7.5.10}$$

At long times when the external momentum k crosses the Hubble radius, this expression simplifies a few-efolds after crossing since in this limit $|S(k;\eta)|^2 \to \mathcal{T}(k)\,(-\pi\eta/4)\,Y_\nu^2(-k\eta)$ and using the expression (7.3.43) we find in this limit

$$\Gamma(k;\eta) \simeq \frac{9\lambda^2\mathcal{T}(0)\mathcal{T}^2(k)}{\pi^2 H^2\Delta(-\eta)(-k\eta)^6} \tag{7.5.11}$$

The Bunch-Davies result is obtained by replacing $\mathcal{T}(k) \to 1$ and coincides with the result obtained in ref.[175][2].

[2] There is a factor 2 error in the prefactor in this reference.

Simple rules:

The analysis presented above yields as corollary the following set of simple rules to assess the effect of non-Bunch-Davies in the correlators:

- Correlation functions feature products of mode functions of the form $S(k,\eta)S^*(k,\eta')$, for values of k so that $-k\eta, -k\eta' \gg 1$ this product can be replaced by

$$S(k,\eta)S^*(k,\eta') \to \frac{\pi}{4}\,\mathcal{T}(k)\left(\eta\,\eta'\right)^{1/2} Y_\nu(-k\eta)Y_\nu(-k\eta')\,. \tag{7.5.12}$$

- In the momentum integrals that lead to infrared divergences and resulting in poles in Δ, the initial condition transfer function can be expanded as $\mathcal{T}(k) \simeq \mathcal{T}(0) + \mathcal{O}(k^2) + \cdots$, the higher order powers of k do not yield infrared enhancements, therefore the poles in Δ are multiplied by $\mathcal{T}(0)$. Namely for poles in Δ that arise from momentum integration it follows that

$$\frac{1}{\Delta} \to \frac{\mathcal{T}(0)}{\Delta}\,. \tag{7.5.13}$$

These simple rules allow to extract the contribution from non-Bunch-Davies initial conditions, encoded in $\mathcal{T}$ to the various correlation functions.

7.6 ENTANGLEMENT ENTROPY: EFFECT OF INITIAL CONDITIONS ON CORRELATIONS ACROSS THE HORIZON

In the $\lambda\phi^3$ theory considered here, a single particle state, $|1_{\vec{k}}\rangle$, decays into a two particle state, $|1_{\vec{k}-\vec{p}}\rangle|1_{\vec{p}}\rangle$ with the corresponding amplitude given by (7.5.2). Full quantum state obtained from the time evolution is a linear superposition of the two particle states summed over the momentum $\vec{p}$. Such a quantum state is *entangled*. This is a general result highlighted in ref.[198]: the decay of a single particle state leads to a *quantum entangled state* with correlations between the daughter particles as a consequence of conservation laws. In a

spatially flat Friedmann-Robertson-Walker cosmology spatial momentum is conserved. In ref.[299] it was realized that the decay of an initial single particle state with wavelength deep inside the Hubble radius produces two particle states which in the case of light fields the leading contribution in Δ corresponds to the decay into a subhorizon and a superhorizon particle. This is an entangled state with correlations between the daughter particles *across the Hubble radius*. As discussed in detail in ref.[299] this process is dominated by the emission and absorption of superhorizon quanta, and therefore it is enhanced in the infrared by poles in Δ which is a hallmark of the infrared aspects associated with light fields in de Sitter (or near de Sitter) space time.

The main tool to study the time evolution of single particle states and the correlated quantum state resulting from the decay is the quantum field theory version of the Weisskopf-Wigner method introduced in refs.[175, 299, 198] where the reader is referred to for a detailed treatment, a brief description is included in appendix (7.8.2) for consistency.

Considering an initial state $|1_{\vec{k}}\rangle$ at initial time η_0 results in the following quantum state

$$|\Psi(\eta)\rangle_I = C_k(\eta)|1_{\vec{k}}\rangle + \sum_{\vec{p}} C_p(k,\eta)|1_{\vec{k}-\vec{p}}\rangle|1_{\vec{p}}\rangle \tag{7.6.1}$$

where the coefficients C_k, C_p are obtained through (7.8.20) and (7.8.26). It has been shown that the Wigner Weisskopf truncation is fully consistent with unitarity as shown in ref. [299]. For completeness, this is shown explicitly in appendix 7.8.2.

With a fully unitary prescription to obtain the coefficients, the pure state density matrix corresponding to the entangled state of eq.(7.6.1) may be written

$$\rho(\eta) = |\Psi(\eta)\rangle\langle\Psi(\eta)|. \tag{7.6.2}$$

Considering the situation where a subhorizon mode ($\vec{k} \gtrsim (-1/\eta)$) decays, tracing out superhorizon ($\vec{p} \lesssim (-1/\eta)$) modes leads to the *mixed state* density matrix for modes whose wavelengths are *inside* the horizon during the evolution. This is given by

$$\rho_r(\eta) = |C_k(\eta)|^2|1_{\vec{k}}\rangle\langle 1_{\vec{k}}| + 2\sum_{p\lesssim(-1/\eta)} |C_p(k;\eta)|^2|1_{\vec{k}-\vec{p}}\rangle\langle 1_{\vec{k}-\vec{p}}| \tag{7.6.3}$$

where the factor 2 accounts for the two regions of superhorizon momenta $p < (-1/\eta)$ and $|\vec{k}-\vec{p}| < (-1/\eta)$ which yield the same contribution, as can be easily seen after a relabelling of momenta.

The entanglement entropy is given by the Von-Neumann entropy for the reduced density matrix, where one finds

$$\mathcal{S}(\eta) = -n_k(\eta)\ln n_k(\eta) - 2\sum_{p\lesssim(-1/\eta)} n_p(\eta)\ln n_p(\eta) \tag{7.6.4}$$

where the occupation numbers of the initial and *produced* quanta are given by

$$n_k(\eta) = \langle\Psi(\eta)|a^\dagger_{\vec{k}}\,a_{\vec{k}}|\Psi(\eta)\rangle = |C_k(\eta)|^2,\;\; n_p(\eta) = \langle\Psi(\eta)|a^\dagger_{\vec{p}}\,a_{\vec{p}}|\Psi(\eta)\rangle = |C_p(k;\eta)|^2\,. \tag{7.6.5}$$

The unitarity relation from eq.(7.8.33) implies that

$$\sum_{\vec{p}} n_p(\eta) = 1 - n_k(\eta)\,. \tag{7.6.6}$$

as expected on physical grounds. At this point, all that remains to calculate the entropy for this process is a calculation of the coefficients, (7.8.20) and (7.8.26).

Using (7.5.4), the coefficient 7.8.26 can be calculated. For $|\vec{p}| \ll -1/\eta; |\vec{k}|, |\vec{k}-\vec{p}| \gg -1/\eta$, the mode functions in 7.4.8 reduce to

$$S_\nu(k,\eta) \to \frac{1}{\sqrt{2k}}\Big[A_k\,e^{-ik\eta} + B_k\,e^{ik\eta}\Big] \;\;;\;\; S_\nu(p,\eta) \to \frac{i}{\sqrt{2}}\frac{A_p - B_p}{(-\eta)^{1-\Delta}p^{3/2-\Delta}} \tag{7.6.7}$$

For momenta k deep inside the Hubble radius the results (7.3.69,7.3.70,7.3.71) justify to set $A_k = 1\;;\; B_k = 0$ to leading order. The integral in (7.5.4) can be carried out with an infrared cutoff $\mu \lesssim (-1/\eta)$ and the leading order in Δ is extracted by approximating $\mathcal{T}(p) \simeq \mathcal{T}(0)$, leading to the result

$$\Sigma(k,\eta_1,\eta_2) = \frac{\alpha}{k^{2-2\Delta}\eta_1^{2-\Delta}\eta_2^{2-\Delta}} \;\;; \tag{7.6.8}$$

where

$$\alpha \equiv \frac{9\lambda^2 \mathcal{T}(0)}{8\pi^2 H^2 \Delta}\,. \tag{7.6.9}$$

Using this result, the coefficient of 7.8.26 becomes

$$C_k(\eta) = \exp\left[-\frac{\alpha}{2z^{2-2\Delta}}\right] \quad ; \quad z \equiv k\eta \tag{7.6.10}$$

The matrix element for this process is given by

$$\begin{aligned}\mathcal{M}(p;k;\eta) &= \langle 1_{\vec{k}-\vec{p}};\, 1_{\vec{p}}|H_I(\eta)|1_{\vec{k}}\rangle = -\frac{6\lambda}{H\eta\sqrt{V}}\, S_\nu(k;\eta)\, S^*_\nu(p;\eta)\, S^*_\nu(|\vec{k}-\vec{p}|;\eta) \\ &\to \frac{-6\lambda(A_0^* - B_0^*)}{2\sqrt{2}kHV^{1/2}(-\eta)^{2-\Delta}p^{3/2-\Delta}}\end{aligned} \tag{7.6.11}$$

so that

$$C_p(k;\eta) = -i\int_{\eta_0}^{\eta} \mathcal{M}(p;k;\eta')\, C_k(\eta')\, d\eta' = \frac{-6i\lambda(A_0^* - B_0^*)}{2\sqrt{2}HV^{1/2}p^{3/2-\Delta}}\frac{1}{\sqrt{\alpha}}\int_{y_0}^{y} e^{-y^2/2}dy \tag{7.6.12}$$

where a change of variables, $\eta = \sqrt{\alpha}/ky$, has been made. In principle, this can be calculated in terms of error functions but unitarity provides a simpler means of evaluation. Since $\alpha \propto |A_0 - B_0|^2/\Delta$, $|C_p(k;\eta)|^2$ can be rewritten as

$$|C_p(k;\eta)|^2 = \frac{\Delta}{Vp^{3-2\Delta}}\frac{|A_0^* - B_0^*|^2}{|A_0 - B_0|^2}F[k,\eta] = \frac{\Delta}{Vp^{3-2\Delta}}F[k,\eta] \tag{7.6.13}$$

The dependence on Δ is a manifestation of unitarity to leading order; if the integral in eq.(7.6.13) is calculated over superhorizon modes, then

$$\sum_{p\lesssim(-1/\eta)} |C_p(k;\eta)|^2 = \frac{F[k;\eta]\,\Delta}{2\pi^2}\int_0^{(-1/\eta)}\frac{p^2dp}{p^{3-2\Delta}} = \frac{F[k;\eta]}{4\pi^2}\,(-1/\eta)^{2\Delta}, \tag{7.6.14}$$

Noting that the Δ in the numerator in eq.(7.6.13) cancels the single pole in Δ from the integral giving an $\mathcal{O}(1)$ contribution, which is what is necessary to satisfy the unitarity condition (7.8.33) to leading order in Δ.

This result is similar to that found in the case of particle decay in Minkowski space time[198]: in this case the particles produced from the decay of a parent particle feature a Lorentzian distribution in energy, with width Γ the decay width of the parent particle and amplitude $1/\Gamma$, so that the energy integral over the distribution is $\mathcal{O}(1)$. In ref.[299] it is proven to leading order in the perturbative expansion $\mathcal{O}(\Gamma)$ that this narrow distribution of large amplitude is the main reason for the fulfillment of unitarity to leading order in the Wigner-Weisskopf approximation. In the case of de Sitter space time, the distribution function of the particles produced with superhorizon wavevectors is $\propto \Delta/p^{3-2\Delta}$ whose momentum integral over the region of superhorizon momenta is also of $\mathcal{O}(1)$.

Thus in the limit $\Delta \ll 1$ the sum $\sum_p |C_p(\eta)|^2$ is dominated by the superhorizon momenta and from the unitarity relation (7.8.33) it is found that

$$\mathrm{Tr}\rho_r(\eta) = |C_k(\eta)|^2 + \sum_p |C_p(\eta)|^2 = 1\,. \tag{7.6.15}$$

To leading order in Δ, the sum is dominated by the superhorizon contributions from both regions of integrations $p \lesssim (-1/\eta)$, $|\vec{k}-\vec{p}| \lesssim (-1/\eta)$ contributing equally, hence

$$\sum_{p\lesssim(-1/\eta)} |C_p(k;\eta)|^2 \simeq \frac{1}{2}\big[1-|C_k(\eta)|^2\big]\,. \tag{7.6.16}$$

Then the factorized form (7.6.13) for superhorizon modes, combined with eqn. (7.6.16) leads to

$$F[k;\eta] = \frac{2\pi^2}{(-\eta)^{-2\Delta}}\left[1-|C_k(\eta)|^2\right], \tag{7.6.17}$$

and for $-k\eta \gg 1$ and $-p\eta \ll 1$ to leading order in Δ, it is found that

$$|C_p(k;\eta)|^2 = \frac{2\pi^2\,\Delta}{V\,p^3\,(-p\eta)^{-2\Delta}}\left[1-|C_k(\eta)|^2\right]; \tag{7.6.18}$$

the same result is valid in the region $-k\eta \gg 1$ with $-|\vec{k}-\vec{p}|\eta \ll 1$ by replacing $p \leftrightarrow |\vec{k}-\vec{p}|$.

The long wavelength limit of eq.(7.6.18) requires a careful treatment. Since $|C_p(\eta)|^2 = n_p(\eta)$ is the distribution function of particles, for a fixed volume V there is an infrared divergence in the occupation as $p \to 0$. However, physically the longest allowed wavelength

must be determined by the linear size of the quantization volume, this forces an introduction of an infrared cutoff:

$$p_m = 1/V^{\frac{1}{3}} \,. \tag{7.6.19}$$

This treatment is similar to the case of Bose-Einstein condensation where momentum integrals are cut off in the infrared with a typical momentum $p_m \propto L^{-1}$ with L being the typical size of the system. At the end of the calculation of thermodynamic variables one takes $L \to \infty$ with a careful analysis of the infrared behavior; the remainder of this calculation proceed in much the same manner.

The definition of the lower momentum cutoff p_m may differ from eq.(7.6.19) by overall constants of $\mathcal{O}(1)$; however, as will be shown in detail in the analysis that follows, this proportionality constant would yield an irrelevant contribution in the limit $\Delta \ll 1$.

Now the calculation of the entanglement entropy is straightforward: Consider

$$I = \sum_{p \leq (-1/\eta)} |C_p(k;\eta)|^2 \ln \left[|C_p(k;\eta)|^2 \right] \equiv I_1 + I_2 \tag{7.6.20}$$

with

$$\begin{aligned} I_1 &= \left[1 - |C_k(\eta)|^2\right] \ln \left[2\pi^2 \Delta \left[1 - |C_k(\eta)|^2\right]\right] \Delta \int_{p_m}^{(-1/\eta)} (-p\eta)^{2\Delta} \frac{dp}{p} \\ &= \frac{1}{2} \left[1 - |C_k(\eta)|^2\right] \ln \left[2\pi^2 \Delta \left[1 - |C_k(\eta)|^2\right]\right] \left[1 - x_m^{2\Delta}\right] \end{aligned} \tag{7.6.21}$$

where the following definition has been made.

$$x_m = (-p_m \eta) \tag{7.6.22}$$

Evaluating I_2 can be done by changing integration variables to $x = -p\eta$ which produces

$$
\begin{aligned}
I_2 &= -\Big[1-|C_k(\eta)|^2\Big]\,\Delta\int_{x_m}^{1} x^{2\Delta-1}\ln\Big[\frac{x^{3-2\Delta}}{x_m^3}\Big]\,dx \\
&= \frac{1}{2}\Big[1-|C_k(\eta)|^2\Big]\Bigg\{\frac{3-2\Delta}{2\Delta}\Big[1-(x_m)^{2\Delta}\Big]+ \qquad (7.6.23)\\
&\quad 3\ln[x_m]\Big[1-(x_m)^{2\Delta}\Big]+(3-2\Delta)\Big[1-(x_m)^{2\Delta}\Big]\ln[x_m]\Bigg\}. \qquad (7.6.24)
\end{aligned}
$$

It is now clear that the limit $x_m \to 0$ may be carried out safely safely in I_1 and in the terms that *do not feature poles in* Δ in I_2. The terms in I_2 that feature the $\ln[x_m]$ and the (single) pole in Δ, namely $(3/2\Delta)\times[1-(x_m)^{2\Delta}]$ yield the leading contribution for $\Delta, x_m \ll 1$.

Therefore for $\Delta \ll 1$ and $x_m \ll 1$, to leading order, the entanglement entropy is found to be

$$
\begin{aligned}
\mathcal{S}(\eta) &\simeq \frac{\alpha}{(k\eta)^2}\,e^{-\frac{\alpha}{(k\eta)^2}}-\Big[1-e^{-\frac{\alpha}{(k\eta)^2}}\Big]\,\ln\Big[1-e^{-\frac{\alpha}{(k\eta)^2}}\Big] \\
&+ \Big[1-e^{-\frac{\alpha}{(k\eta)^2}}\Big]\Bigg\{\ln\Big[\frac{1}{2\pi^2\,\Delta}\Big]+\frac{3}{2\Delta}\Big[W[\eta]-1+e^{-W[\eta]}\Big]+\mathcal{O}(\Delta)\Big]\Bigg\} \qquad (7.6.25)
\end{aligned}
$$

where

$$
W[\eta]=\frac{2\Delta}{3}\,\ln\Big[V_{ph}(\eta)\,H^3\Big]\quad;\quad V_{ph}(\eta)=V\,(C(\eta))^3\,, \qquad (7.6.26)
$$

with $C(\eta)=a(t(\eta))$ is the scale factor and α is given in eqn. (7.6.9). The function $W[\eta]-1+e^{-W[\eta]}$ is manifestly (semi) positive and monotonically increasing, behaving as $\simeq W^2/2$ for $W \ll 1$ and as $\simeq W$ for $W \gg 1$. As $\eta \to 0$ the entanglement entropy grows monotonically with the physical volume.

An important consequence of unitarity is that the dependence of the entanglement entropy on the initial conditions is only through α.

The logarithmic volume dependence is similar to the result obtained in Minkowski space time, and its interpretation is that asymptotically the entropy saturates to the logarithm of the number of accessible states in phase space, which is proportional to the volume. However in the expanding cosmology it is the physical volume that enters in the final expression; as

the cosmological expansion proceeds the available phase space increases as more and more wavevectors cross the Hubble radius. Furthermore the infrared enhancement from light fields during inflation translate in the $\ln[\Delta]$. It is clear from the expression above that the definition of p_m in (7.6.19) differed by a proportionality constant $\mathcal{C} \simeq \mathrm{O}(1)$, the expression above would have been modified by an term $\sim \Delta \ln[\mathcal{C}] \ll 1$ which can be safely neglected, thus confirming that the choice of the minimal value of the momentum (infrared cutoff) (7.6.19) is insensitive to multiplicative factors of $\mathcal{O}(1)$ for $\Delta \ll 1$.

7.7 CONCLUSIONS AND FURTHER QUESTIONS

The recent CMB data from Planck distinctly shows a persistence of large scale anomalies, among them a suppression of the power spectrum for large scales, in the region of the Sachs-Wolfe plateau for $l \lesssim 10$. Motivated by the possibility that these anomalies, in particular the suppression of power at low multipoles, is of primordial origin perhaps heralding new physics on superhorizon scales, we studied the effect of initial conditions arising from a rapid evolution of the inflaton during a brief stage prior to slow roll. Such a rapid evolution, or "fast roll" stage leads to the equations for the mode functions of scalar and tensor perturbations that features a potential which is localized in conformal time. The effect of this potential translates into non-Bunch-Davies conditions on the mode functions during the slow roll stage, the Bogoliubov coefficients being determined by the properties of the potential during the pre-slow roll stage.

Implementing methods from potential scattering theory we obtained general properties of these Bogoliubov coefficients, in particular their superhorizon and sub-horizon behavior. The effect of these initial conditions on the power spectra of scalar and tensor perturbations are encoded in an initial condition transfer function $\mathcal{T}(k)$. We showed that for wavevectors that exited the Hubble radius during the very early stages of slow roll the large scale transfer function $\mathcal{T}(k \approx 0)$ leads to a suppression of the power spectrum for attractive potentials, such as those found previously for the case of a "fast-roll stage"[275, 276, 277]. Furthermore for modes that are inside the Hubble radius during most of the slow roll stage $\mathcal{T}(k) \lesssim 1/k^4$

suggesting that the effect of initial conditions determined by pre-slow roll stage is strongly suppressed for higher multipoles and would not modify the small scale aspects of the CMB, such as acoustic peaks.

Since the initial conditions impact mainly large scales, we were motivated to study their effect on the infrared sector of typical minimally coupled scalar field theories with typical self-interactions $\lambda\phi^p$ with $p = 3, 4$ when the slow roll stage is a (nearly) de Sitter cosmology. The correlation functions of light scalar fields with mass $M \ll H$ (H is the Hubble parameter during de Sitter inflation), feature infrared divergences manifest as poles in $\Delta = M^2/3H^2$. These infrared divergences lead to a dynamical generation of mass if the bare mass of the scalar field vanishes.

For $p = 3$ we find that the infrared singularity of bare massless theory leads to the formation of a non-perturbative condensate which reaches a fixed value at long times and implies the dynamical generation of a mass $M = \sqrt{3}\,H\,\left(\frac{\lambda}{2\pi H}\right)^{1/3}\left[\mathcal{T}(0)\right]^{1/6}$. For $p = 4$ we find $M = H\left[\frac{9\lambda\,\mathcal{T}(0)}{2\pi^2}\right]^{1/4}$. In both cases the emergence of a dynamical infrared generated mass yields scalar power spectra with *anomalous dimensions* that depend non-analytically on initial conditions, namely $P_s(k) \propto k^{\Delta}$ where for $p = 3, 4$ respectively we find

$$\Delta_{(3)} = \left[\frac{\lambda\,\sqrt{\mathcal{T}(0)}}{2\pi\,H}\right]^{\frac{2}{3}} \quad ; \quad \Delta_{(4)} = \left[\frac{\lambda\,\mathcal{T}(0)}{2\pi^2}\right]^{\frac{1}{2}}. \tag{7.7.1}$$

In an expanding cosmology all the quanta of a field can decay into quanta of the *same field* as a consequence of the lack of energy conservation and kinematic thresholds. The time dependent decay width of single particle states are enhanced by the infrared divergences that are also responsible for the dynamical generation of mass. We obtain the modification of the decay width for single particle states induced by the non-Bunch -Davies initial conditions, for $p = 3$ we find

$$\Gamma(k;\eta) \simeq \frac{9\lambda^2\mathcal{T}(0)\mathcal{T}^2(k)}{\pi^2 H^2\Delta(-\eta)(-k\eta)^6} \tag{7.7.2}$$

The decay of a single particle state yields an *entangled* quantum state of the daughter particles, entanglement being a consequence of momentum conservation. We implement field theoretical version of the Wigner-Weisskopf method adapted to inflationary cosmology to obtain the full quantum state that results from the time evolution and decay of an initial

single particle state. This method yields manifestly unitary time evolution of the quantum state. In ref.[299] it was realized that this quantum state features entanglement and correlations between sub and superhorizon quanta. Tracing over the superhorizon degrees of freedom leads to an entanglement entropy that grows as more modes exit the horizon during inflation. We obtain the modifications of this entanglement entropy from non-Bunch-Davies initial conditions. These affect only the decay width of the parent particle and is only through this (time dependent) width that the entanglement entropy.

In all cases studied in this article, the initial conditions from a "fast roll" stage prior to slow roll that result in an initial condition transfer function that suppresses the power of scalar perturbations at large scales, also result in a suppression of the infrared effects: dynamical masses, anomalous dimensions of scalar power spectra and decay widths of quantum states.

The next stage of the study of the impact of initial conditions from a pre-slow roll stage will focus on the modifications on the curvature and gravitational wave power spectra for different inflationary scenarios that are supported by the WMAP[246, 210, 249] and Planck[63] data. This stage will necessarily involve a numerical analysis of the potentials from a fast roll stage along with a study of possible initial conditions on the velocity of the inflaton field. We will report on current work on these issues in a forthcoming study.

7.8 APPENDICES

7.8.1 Calculation of $\Sigma(k;\eta_1,\eta_2)$

In this appendix we calculate the self-energy (7.5.4) to leading order in Δ and in the long time limit $\eta_1,\eta_2 \to 0$.

The first step is to perform the angular integration in (7.5.4). Making the substitution

$p \equiv |k-q| = \sqrt{k^2+q^2-2kq\cos\theta}$ and $d(\cos\theta) = -p\,dp/kq$ so that

$$\int d^3q f(|q|)g(|k-q|) = 2\pi \int d(\cos\theta) \int dq\ q^2[...] = \frac{2\pi}{k}\int_0^\infty dq\ q\ f(|q|) \int_{|k-q|}^{k+q} dp\ p\ g(|p|) \tag{7.8.1}$$

This simplifies the integration to

$$\begin{aligned}
\Sigma(k,\eta_1,\eta_2) &= \frac{9\lambda^2}{\pi^2 H^2 k\eta_1\eta_2} S^*(k,\eta_1)S(k,\eta_2) \\
&\times \int_0^\infty dq\, qS(q,\eta_1)S^*(q,\eta_2) \int_{|k-q|}^{k+q} dp\, pS(p,\eta_1)S^*(p,\eta_2) \\
&\equiv \frac{9\lambda^2}{\pi^2 H^2 k\eta_1\eta_2} S^*(k,\eta_1)S(k,\eta_2)J(k,\eta_1,\eta_2)
\end{aligned} \tag{7.8.2}$$

where

$$J(k,\eta_1,\eta_2) = \int_0^\infty dq\, q\, S(q,\eta_1)S^*(q,\eta_2) \int_{|k-q|}^{k+q} dp\, p\, S(p,\eta_1)S^*(p,\eta_2) \tag{7.8.3}$$

As with the tadpole, this integral features infrared divergences for massless, minimally coupled fields. From the discussion of the tadpoles, it should be clear that there are infrared divergences for $q, p \to 0$, namely in the integration regions $q \simeq 0$; $q \simeq k$. The integral is evaluated with the same method as for the tadpole, isolating the regions of infrared divergences by introducing an infrared cutoff, keeping the most infrared singular terms of the mode functions in the band of wavevectors up to the infrared cutoff extracting the leading order poles in Δ and set $\nu = 3/2$ for wavevectors larger than the cutoff since these integrals are infrared finite for finite cutoff in the limit $\Delta \to 0$. Therefore, we write in obvious notation

$$J = \int_0^\mu dq\,[...] + \int_\mu^\infty dq\,[...] \equiv J_< + J_> \tag{7.8.4}$$

The $J_<$ integral is evaluated by using $q < \mu \sim 0$ so that with $k \gg \mu$ the argument of the p-integral can be evaluated at $p = k$ and the p integral becomes simply $2kqS(k,\eta_1)S^*(k,\eta_2)$ and

$$\begin{aligned} J_< &= \int_0^\mu dq\, q\, S(q,\eta_1)S^*(q,\eta_2) \int_{|k-q|}^{k+q} dp\, p\, S(p,\eta_1)S^*(p,\eta_2) \\ &\sim 2k\, S(k,\eta_1)\, S^*(k,\eta_2) \int_0^\mu dq\, q^2\, S(q,\eta_1)S^*(q,\eta_2) \end{aligned} \tag{7.8.5}$$

Using the long wavelength and long time form of the mode functions given by eqn. (7.3.63) we find

$$J_< = S(k,\eta_1)\, S^*(k,\eta_2) \left[\left(\frac{4}{\eta_1 \eta_2} \right)^{\nu-1/2} \frac{k\, \Gamma^2(\nu)}{\pi} \right] \mathcal{T}(0)\, \frac{\mu^{2\Delta}}{2\Delta} \tag{7.8.6}$$

To evaluate the $J_>$ integral, care must be taken around the poles. There will be infrared divergences for $q = k$ so that the integral is separated as

$$J_> = \int_\mu^\infty dq[...] = \underbrace{\int_\mu^{k-\mu} dq[...]}_{J_>^{(a)}} + \underbrace{\int_{k-\mu}^{k} dq[...]}_{J_>^{(b)}} + \underbrace{\int_{k}^{k+\mu} dq[...]}_{J_>^{(c)}} + \underbrace{\int_{k+\mu}^{\infty} dq[...]}_{J_>^{(d)}} \tag{7.8.7}$$

Since the integrals away from the infrared limit, namely $J_>^{(a)/(d)}$, are finite for finite μ, we can set in these integrals $\nu = 3/2$ as they do not feature poles in Δ. In which case, these integrals are subleading with respect to Δ and need not be considered for a leading order calculation.

The only integrals remaining for the leading order contribution are $J_>^{(b,c)}$. Consider

$$J_>^{(b)} = \int_{k-\mu}^{k} dq\, q\, S(q,\eta_1)S^*(q,\eta_2) \int_{|k-q|}^{k+q} dp\, p\, S(p,\eta_1)S^*(p,\eta_2)\,, \tag{7.8.8}$$

after the change of variable $q = k - r$, to leading order we obtain

$$J_>^{(b)} \simeq k\, S(k,\eta_1)\, S^*(k,\eta_2) \int_0^\mu dr \int_r^{2k+r} dp\, p\, S(p,\eta_1)S^*(p,\eta_2)\,, \tag{7.8.9}$$

the leading order contribution arises from the lower limit of the r integral, this contribution is obtained by integrating in a small region around the lower limit using the mode functions (7.3.63) and approximating $\mathcal{T}(p) \simeq \mathcal{T}(0) + \mathcal{O}(p^2) + \cdots$ and keeping only the $p = 0$ term in this expansion because the higher order terms will not yield poles in Δ, we find

$$\int_r^{2k+r} dp\, p\, S(p,\eta_1) S^*(p,\eta_2) = \frac{\Gamma(\nu)\Gamma(\nu-1)}{2\pi^2} \left(\frac{4}{\eta_1\eta_2}\right)^{\nu} \frac{\pi}{4} \Big(\eta_1\,\eta_2\Big)^{1/2} \mathcal{T}(0)\; r^{2-2\nu} + \cdots \tag{7.8.10}$$

where the dots stand for terms that will not yield poles in Δ as $\Delta \to 0$. Finally carrying out the r-integral we find

$$J_>^{(b)} = k\, S(k,\eta_1)\, S^*(k,\eta_2) \frac{\Gamma(\nu)\Gamma(\nu-1)}{2\pi^2} \left(\frac{4}{\eta_1\eta_2}\right)^{\nu} \frac{\pi}{4} \Big(\eta_1\,\eta_2\Big)^{1/2} \frac{\mu^{2\Delta}}{2\Delta} + \cdots . \tag{7.8.11}$$

The next term $J_>^{(c)}$ can be evaluated in a similar manner, but now changing variables in the q-integral to $q = k + r$ and recognizing that the lower limit in the p-integral is now $q - k = r$ upon changing variables in the q-integral. Again the p integral is dominated by the lower limit which can be extracted just as in the previous case finally leading to

$$J_>^{(c)} = J_>^{(b)} . \tag{7.8.12}$$

Expanding the pole terms

$$\frac{\mu^{2\Delta}}{2\Delta} = \frac{1}{2\Delta} + \ln[\mu] + \cdots \tag{7.8.13}$$

all the terms with $\ln[\mu]$ will cancel among all the different contributions, this is easily seen by taking the μ derivative of J given by eqn. (7.8.4) as the arbitrary cutoff μ has been introduced simply to split the integrals and the total integral cannot depend on μ.

Finally, to leading order

$$J = J_< + J_>^{(b)} + J_>^{(c)} + \mathcal{O}(\Delta^0) = 2k \frac{S(k,\eta_1) S^*(k,\eta_2)}{\eta_1\eta_2} \mathcal{T}(0) \left(\frac{1}{\Delta} + \mathcal{O}(\Delta^0)\right) \tag{7.8.14}$$

Combining this result with (7.8.2) we finally find,

$$\Sigma(k,\eta_1,\eta_2) = \frac{18\,\lambda^2\,\mathcal{T}(0)}{\pi^2\,H^2\,\Delta} \frac{|S(k,\eta_1)|^2}{(\eta_1)^2} \frac{|S(k,\eta_2)|^2}{(\eta_2)^2} + \mathcal{O}(\Delta^0) . \tag{7.8.15}$$

7.8.2 Wigner-Weisskopf theory and unitarity

In this apprendix we summarize the main aspects of the non-perturbative Wigner-Weisskopf method to study the quantum state from particle decay for consistency. More details are available in refs.[175, 299, 198].

The interaction picture states are expanded in terms of Fock states associated with the creation and annihilation operators $\alpha_k, \alpha_k^\dagger$, namely

$$|\Psi(\eta)\rangle_I = \sum_n C_n(\eta)|n\rangle \tag{7.8.16}$$

As shown in earlier, the time evolution of a state in the interaction picture is given by

$$i\frac{d}{d\eta}|\Psi(\eta)\rangle_I = \hat{H}_I(\eta)|\Psi(\eta)\rangle_I \tag{7.8.17}$$

so that the (conformal) time evolution of the coefficients is given by

$$\frac{d}{d\eta}C_n(\eta) = -i\sum_m C_m(\eta)\langle n|\hat{H}_I(\eta)|m\rangle \tag{7.8.18}$$

While this is exact, the solution is an *infinite hierarchy* and finding an exact solution is impractical. This can be vastly simplified by making the assumption that the initial state, $|A\rangle$, only couples to a single set of intermediate states, $|\kappa\rangle$, where this assumption is exact if the situation is confined to processes of $\mathcal{O}(H_I)$ (which is valid for this work). Under this assumption, the coefficients obey

$$\begin{aligned}\frac{d}{d\eta}C_A(\eta) &= -i\sum_\kappa \langle A|H_I(\eta)|\kappa\rangle C_\kappa(\eta)\\ \frac{d}{dt}C_\kappa(\eta) &= -i\langle\kappa|H_I(\eta)|A\rangle C_A(\eta)\end{aligned} \tag{7.8.19}$$

where $\sum_\kappa$ is over all states that couple to $|A\rangle$ via first order in H_I.

Considering the general situation of particle decay, $A \to \kappa_1, \kappa_2, ...$, where initally at some time, $\eta = \eta_o$, the state is given by $|\Psi(\eta_o)\rangle = |A\rangle$. This is equivalent to the initial condition $C_n(\eta_o) = \delta_{n,A}$. Upon integrating the second of 7.8.19, one obtains

$$\begin{aligned} C_\kappa(\eta) &= -i \int_0^\eta d\eta' \langle \kappa | H_I(\eta') | A \rangle C_A(\eta') \\ \frac{d}{d\eta} C_A(\eta) &= -\int_0^\eta d\eta' \sum_\kappa \langle A | H_I(\eta) | \kappa \rangle \langle \kappa | H_I(\eta') | A \rangle C_A(\eta') \end{aligned} \tag{7.8.20}$$

It proves useful to make the definition

$$\Sigma_A(\eta, \eta') = \sum_\kappa \langle A | H_I(\eta) | \kappa \rangle \langle \kappa | H_I(\eta') | A \rangle \tag{7.8.21}$$

Note that this is equal to 7.5.4. Then

$$\frac{d}{d\eta} C_A(\eta) = -\int_{\eta_o}^\eta d\eta' \Sigma_A(\eta, \eta') C_A(\eta') \tag{7.8.22}$$

The relation between this method and the Dyson resummation is discussed in detail in ref. [229]. It can be shown that this treatment is non-perturbative and the time evolution of the coefficients are slow which justifies a derivative expansion. The derivative expansion is done by introducing the term

$$W_0(\eta, \eta') = \int_{\eta_o}^{\eta'} d\eta'' \Sigma_A(\eta, \eta'') \quad ; \quad \frac{d}{d\eta'} W_0(\eta, \eta') = \Sigma_A(\eta, \eta') \quad ; \quad W_0(\eta, \eta_o) = 0 \tag{7.8.23}$$

So that integrating 7.8.22 by parts leads to

$$\int_{\eta_o}^\eta d\eta' \Sigma_A(\eta, \eta') C_A(\eta') = W_0(\eta, \eta) C_A(\eta) - \int_{\eta_o}^\eta d\eta' W_0(\eta, \eta') \frac{d}{d\eta'} C_A(\eta') \tag{7.8.24}$$

For a weakly interacting theory, such that $H_I \sim \mathcal{O}(\lambda)$ and $\lambda \ll 1$, the second term is at higher order in perturbation theory and may be discarded. To leading order, 7.8.22 simplifies drastically to

$$\frac{d}{d\eta} C_A(\eta) = -W_0(\eta, \eta) C_A(\eta) + \mathcal{O}(\lambda^4) \tag{7.8.25}$$

with the simple solution

$$C_A(\eta) = e^{-\int_{\eta_o}^{\eta} d\eta' W_0(\eta',\eta')} \tag{7.8.26}$$

Interpretation of this result follows from the analysis in Minkowski spacetime. It has been shown that the imaginary part of the integral will provide the second order energy shift while the real part provides the decay width, similar to Fermi's golden rule. This is made explict in the literature with the result that

$$\int_{\eta_o}^{\eta'} d\eta'' \Sigma_A(\eta', \eta'') = i\delta E_A^{(1)}(\eta') + \frac{1}{2}\Gamma(\eta') \tag{7.8.27}$$

Where the real part matches 7.5.7 exactly. Finally, the full time dependence of the coefficient can be written as

$$C_A(\eta) = e^{-i\int d\eta' \delta E^{(1)}(\eta')} e^{-\frac{1}{2}\int d\eta' \Gamma_A(\eta')} \tag{7.8.28}$$

Since the probability of measuring particle A is $|C_A|^2$ and with the discussion in section 7.5.1, the interpretation of Γ as the decay rate is clear. It has also been shown that the Wigner Weisskopf method produces the same results for the self consistent mass generation discussed earlier [229].

One the main goals is to study the entanglement entropy from tracing over superhorizon degrees of freedom. Thus it is important to make sure that the loss of information encoded in the entanglement entropy is a genuine effect of the tracing procedure and not a consequence of approximations in the evolution of the quantum state. In this appendix, the discussion follows ref. [198, 299] where it is shown that the Wigner-Weisskopf approximation and its Markovian implementation maintain unitary time evolution.

Using (7.8.20) consider

$$\sum_{\kappa} |C_\kappa(\eta)|^2 = \int_{\eta_0}^{\eta} d\eta_1 C_A^*(\eta_1) \int_{\eta_0}^{\eta} d\eta_2 \Sigma(\eta_1, \eta_2) C_A(\eta_2). \tag{7.8.29}$$

Inserting $1 = \Theta(\eta_1 - \eta_2) + \Theta(\eta_2 - \eta_1)$, it follows that

$$\sum_{\kappa} |C_\kappa(\eta)|^2 = \int_{\eta_0}^{\eta} d\eta_1 C_A^*(\eta_1) \int_{\eta_0}^{\eta_1} d\eta_2 \Sigma(\eta_1,\eta_2) C_A(\eta_2) + \int_{\eta_0}^{\eta} d\eta_2 C_A(\eta_2) \int_{\eta_0}^{\eta_2} d\eta_1 \Sigma(\eta_1,\eta_2) C_A^*(\eta_1). \tag{7.8.30}$$

Using $\Sigma(\eta_1,\eta_2) = \Sigma^*(\eta_2,\eta_1)$, relabelling $\eta_1 \leftrightarrow \eta_2$ in the second line of (7.8.30) and using (7.8.22), one can show

$$\begin{aligned} \sum_{\kappa} |C_\kappa(\eta)|^2 &= -\int_{\eta_0}^{\eta} d\eta_1 \Big[C_A^*(\eta_1) \frac{d}{d\eta_1} C_A(\eta_1) + C_A(\eta_1) \frac{d}{d\eta_1} C_A^*(\eta_1) \Big] \\ &= -\int_{\eta_0}^{\eta} d\eta_1 \frac{d}{d\eta_1} |C_A(\eta_1)|^2 = 1 - |C_A(\eta)|^2 \end{aligned} \tag{7.8.31}$$

where the initial condition $C_A(\eta_0) = 1$ has been used. This is the statement of unitary time evolution, namely

$$|C_A(\eta)|^2 + \sum_{\kappa} |C_\kappa(\eta)|^2 = |C_A(\eta_0)|^2 \tag{7.8.32}$$

To leading order in the Markovian approximation, the unitarity relation becomes

$$\sum_{\kappa} |C_\kappa(\eta)|^2 = -2 \int_{\eta_0}^{\eta} \Big| C_A(\eta_1) \Big|^2 \, \mathrm{Re}\Big[W_0(\eta_1,\eta_1) \Big] \, d\eta_1 = 1 - |C_A(\eta)|^2 \tag{7.8.33}$$

where $C_A(\eta_0) = 1$.

8.0 TENSOR TO SCALAR RATIO AND LARGE SCALE POWER SUPPRESSION FROM PRE-SLOW ROLL INITIAL CONDITIONS.

Based on: (ref. [283])

L. Lello, D. Boyanovsky, JCAP 05 02 (2014)

8.1 INTRODUCTION

Inflation not only provides a solution to the horizon and flatness problems but also furnishes a mechanism for generating scalar (curvature) and tensor (gravitational wave) quantum fluctuations[237, 238, 239, 240]. These fluctuations seed the small temperature inhomogeneities in the cosmic microwave background (CMB) upon reentering the particle horizon during recombination. Most inflationary scenarios predict a nearly gaussian and nearly scale invariant power spectrum of adiabatic fluctuations [241, 245, 242, 243, 244]. These important predictions of inflationary cosmology are supported by observations of the cosmic microwave background[246, 210, 63, 300] which are beginning to discriminate among different scenarios.

Recent results from the Planck collaboration[63, 300, 301] have provided the most precise analysis of the (CMB) to date, confirming the main features of the inflationary paradigm, but at the same time highlighting perplexing large scale anomalies, some of them, such as a low quadrupole, dating back to the early observations of the Cosmic Background Explorer (COBE)[247, 248], confirmed with greater accuracy by WMAP[249] and Planck[63, 300, 301]. Recently the BICEP2 collaboration [302] has provided the first measurement of primordial

B-waves, possibly the first direct evidence of inflation.

The interpretation and statistical significance of these anomalies is a matter of much debate, but being associated with the largest scales, hence the most primordial aspects of the power spectrum, their observational evidence is not completely dismissed[252]. The possible origin of the large scale anomalies is vigorously discussed, whether these are of primordial origin or a consequence of the statistical analysis (masking) or secondary anisotropies is still an open question. Some studies claim the removal of large scale anomalies (including the suppression of power of the low multipoles) after substraction of the integrated Sachs-Wolfe effect (ISW)[250], however a different recent analysis[251] finds that the low quadrupole becomes even *more* anomalous after subtraction of the (ISW) contribution, although some expansion histories may lead to an (ISW) suppression of the power spectrum[303]. The most recent Planck[63, 300, 301] data still finds a statistically significant discrepancy at low multipoles, reporting a power deficit $5-10\%$ at $l \lesssim 40$ with $2.5-3\,\sigma$ significance. This puzzling and persistent result stands out in an otherwise consistent picture of ΛCDM insofar as the (CMB) power spectrum is concerned. Recent analysis of this lack of power at low l[211] and large angles[252], suggests that while limited by cosmic variance, the possibility of the primordial origin of the large scale anomalies cannot be dismissed and merits further study.

The simpler inflationary paradigm that successfully explains the cosmological data relies on the dynamics of a scalar field, the inflaton, evolving slowly during the inflationary stage with the dynamics determined by a fairly flat potential. This simple, yet observationally supported inflationary scenario is referred to as slow-roll inflation[245, 241, 242, 243, 244]. Within this scenario wave vectors of cosmological relevance cross the Hubble radius during inflation with nearly constant amplitude leading to a nearly scale invariant power spectrum. The quantization of the gaussian fluctuations (curvature and tensor) is carried out by imposing a set of initial conditions so that fluctuations with wavevectors deep inside the Hubble radius are described by Minkowski space-time free field mode functions. These are known as Bunch-Davies initial conditions[253] (see for example[245, 241, 243, 244] and references therein).

The issue of modifications of these initial conditions and the potential impact on the inflationary power spectra[254, 255, 256, 257, 258, 259, 260, 261, 262, 208, 304], enhancements to non-gaussianity[265, 266, 267, 268, 269, 270, 271, 280, 281, 282, 305], and large scale structure[272] have been discussed in the literature. Furthermore, arguments presented in refs.[208, 188] suggest that Bunch-Davies initial conditions are not the most natural ones to consider and may be unstable to small perturbations.

Whereas the recent results from Planck[63, 300, 301] provide tight constraints on primordial non-gaussianities including modifications from initial conditions, these constraints *per se* do not apply directly to the issue of initial conditions on other observational aspects.

Non-Bunch-Davies initial conditions arising from a pre-slow roll stage during which the (single) inflaton field features "*fast-roll*" dynamics have been proposed as a *possible* explanation of power suppression at large scales[273, 274, 275, 276, 306, 277, 279]. This kinetically dominated fast roll stage has been considered previously using the Hamilton-Jacobi form of the Friedman equations in ref.[307] with some of the consequences due to this type of scalar driven cosmology examined in detail in[308, 309, 310]. Alternative pre-slow-roll descriptions in terms of interpolating scale factors pre (and post) inflation have also been discussed in ref.[278]. The influence of non-Bunch Davies initial conditions arising from a fast-roll stage just prior to slow roll on the infrared aspects of nearly massless scalar fields in de Sitter space time have been studied in ref.[311]. Recent work [312] has shown that a kinetically dominated regime is in fact quite a generic feature under a very broad class of single field inflationary models providing further incentive for consideration of fast roll scenarios.

Motivations, goals and summary of results:

Inflationary scenarios predict the generation of primordial gravitational waves and their detection remains one of the very important goals of observational cosmology. Planck[63] has placed constraints on the tensor to scalar ratio of $r < 0.11\,(95\%CL)$ while the BICEP [302] experiment has recently reported a measurement of $r = 0.20^{+0.07}_{-0.05}$. The BICEP value is much larger than many had expected and there exist models which can generate enhancements, refs [313] for example, which could explain the largeness of this value.

Suggestions of how to relieve the tension between the two experiments have been put

forth where a possible solution invokes a running of the spectral index. Recently, in ref.[314], a comparison between models featuring a running spectral index and models with a large scale power suppression has been made with the aim of determining which model relieved the tension most effectively. It was shown in this reference that a large scale power suppression of 35% yielded a considerably better fit to data than allowing a running of the spectral index, further improving the claims that the low l anomaly should be taken seriously.

The high amount of tension between these two experiments may be alleviated by future and forthcoming observations that will continue to constrain this important quantity, a quantity from which ultimately the scale of inflation may be extracted[63],

$$V_{inf} = (1.94 \times 10^{16}\,\mathrm{GeV})^4 \left(\frac{r}{0.12}\right). \tag{8.1.1}$$

A distinct prediction of single field slow-roll inflationary models with a standard kinetic term is

$$r = 16\epsilon_V = -8n_T \tag{8.1.2}$$

with ϵ_V a (potential) slow roll parameter and n_T is the index of the power spectrum of gravitational waves. The relation (8.1.2) is often quoted as a "consistency relation". This relation is obtained by imposing Bunch-Davies initial conditions on tensor perturbations during the near de Sitter slow roll stage[245, 241, 243, 244].

Our goals in this article are the following:

- Motivated by the recent results from the PLANCK collaboration[63, 300, 301] reporting the persistence of anomalies for small l and large angular scales we study the modifications to the power spectra of curvature and tensor perturbations and the tensor to scalar ratio arising from non-Bunch Davies initial conditions imprinted from a pre-slow roll stage in which the dynamics of the scalar field is dominated by the kinetic term, namely a "fast-roll stage". While previous studies of modifications of the scalar power spectrum from a fast roll stage focused on specific realizations of the inflationary potential, our goal is to extract the main corrections *without resorting to a specific choice of the potential* but by parametrizing the fast roll stage by the initial ratio of kinetic to potential energy of the inflaton, $\dot{\Phi}_i^2/2V = \kappa$, and the potential slow roll parameters ϵ_V, η_V

which have been constrained by Planck and WMAP-polarization (Planck+WP)[63] to be $\epsilon_V < 0.008\ (95\%\, CL); \eta_V = -0.010^{+0.005}_{-0.011}$.

- To explore possible *correlations* between the suppression of the low multipoles in the temperature power spectrum, and features in the tensor to scalar ratio $r(k_0)$ as a function of the pivot scale, and more generally, to the power spectrum of tensor perturbations, as a consequence of the fast roll stage.

- To assess the scales and general aspects of features in the power spectra resulting from the modification of the initial conditions and their potential observability.

Brief summary of results: A fast roll stage prior to slow roll leads to non-Bunch-Davies conditions on the observationally relevant mode functions that cross the Hubble radius during slow roll. These modifications yield oscillatory corrections to the power spectra of curvature and tensor perturbations, with a period determined by the Hubble scale during slow roll inflation, and a modification of the consistency condition for the tensor to scalar ratio r with oscillatory features as a function of the pivot scale. The results are general and do not depend on the specific form of the inflationary potential but to leading order in slow roll depend only on $\kappa; \epsilon_V; \eta_V$. We describe a systematic program that yields the solution interpolating between the fast and slow roll stages based on a derivative expansion and separation of scales, which is *independent of the inflationary potentials* provided these are monotonic and can be described in a derivative expansion characterized by slow roll parameters. The Non-Bunch Davies initial conditions from the fast roll stage lead to corrections to the power spectra of scalar and tensor perturbations in the form of oscillatory features with a typical frequency determined by the Hubble scale during slow roll. The corrections to the power spectrum for curvature perturbations lead to a suppression of the quadrupole that is correlated with the oscillatory features in the tensor to scalar ratio $r(k_0)$ as a function of the pivot scale k_0. The quadrupole suppression is consistent with the latest results from Planck[300] and the oscillatory features in $r(k_0)$ *could* be observable[315] *if* the mode corresponding to the Hubble radius today crossed the Hubble radius a few e-folds from the beginning of slow roll.

8.2 FAST ROLL STAGE:

We consider a spatially flat Friedmann-Robertson-Walker (FRW) cosmology with

$$ds^2 = dt^2 - a^2(t)(d\vec{x})^2 = C^2(\eta)[d\eta^2 - (d\vec{x})^2] \;\; ; \;\; C(\eta) \equiv a(t(\eta)) \,, \tag{8.2.1}$$

where t and η stand for cosmic and conformal time respectively and consider curvature and tensor perturbations. The dynamics of the scale factor in single field inflation is determined by Friedmann and covariant conservation equations

$$H^2 = \frac{1}{3M_{Pl}^2}\left[\frac{1}{2}\dot{\Phi}^2 + V(\Phi)\right] \;\; ; \;\; \ddot{\Phi} + 3H\dot{\Phi} + V'(\Phi) = 0 \,. \tag{8.2.2}$$

During the slow roll near de Sitter stage,

$$H_{sr}^2 \simeq \frac{V_{sr}(\Phi)}{3M_{Pl}^2} \;\; ; \;\; 3H\dot{\Phi} + V'_{sr}(\Phi) \simeq 0 \,. \tag{8.2.3}$$

This stage is characterized by the smallness of the (potential) slow roll parameters[241, 245, 242, 243, 244]

$$\epsilon_V = \frac{M_{Pl}^2}{2}\left[\frac{V'_{sr}(\Phi)}{V_{sr}(\Phi)}\right]^2 \simeq \frac{\dot{\Phi}_{sr}^2}{2M_{Pl}^2H^2} \quad , \quad \eta_V = M_{Pl}^2\,\frac{V''_{sr}(\Phi)}{V_{sr}(\Phi)} \,, \tag{8.2.4}$$

(here $M_{Pl} = 1/\sqrt{8\pi\, G}$ is the *reduced* Planck mass).

Instead, in this section we consider an initial stage dominated by the kinetic term, namely a fast roll stage, thereby neglecting the term V' in the equation of motion for the inflaton, (8.2.2) and consider the potential to be (nearly) constant and equal to the potential during the slow roll stage, namely $V(\Phi) \simeq V(\Phi_{sr}) \equiv V_{sr}$. In the following section we relax this condition in a consistent expansion in $\sqrt{\epsilon_V}$.

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{1}{3M_{Pl}^2}\left[\frac{1}{2}\dot{\Phi}^2 + V_{sr}\right] \tag{8.2.5}$$

$$\ddot{\Phi} + 3H\dot{\Phi} \simeq 0 \,. \tag{8.2.6}$$

The solution to (8.2.6) is given by

$$\dot{\Phi}(t) = \dot{\Phi}_i\Big(\frac{a_i}{a(t)}\Big)^3 \,, \tag{8.2.7}$$

an initial value of the velocity damps out and the slow roll stage begins when $\ddot{\Phi} \ll 3H_{sr}\dot{\Phi} \simeq -V'_{sr}(\Phi)$. During the slow roll stage when $3H_{sr}\dot{\Phi}_{sr} \simeq -V'_{sr}$ it follows that

$$\frac{3\dot{\Phi}_{sr}^2}{2V_{sr}} = \epsilon_V \,. \tag{8.2.8}$$

The dynamics enters the slow roll stage when $\dot{\Phi} \sim \mathcal{O}(\sqrt{\epsilon_V})$ as seen by (8.2.4). To a first approximation, we will assume that Eq.(8.2.7) holds not only for the kinetically dominated epoch, but also until the beginning of slow roll ($\dot{\Phi}^2 \sim \epsilon_V$). In Section 8.3, this approximation is justified and the error incurred from such an assumption is made explicit. The dynamics enters the slow roll stage at a value of the scale factor $a(t_{sr}) \equiv a_{sr}$ so that

$$\dot{\Phi}_{sr} a_{sr}^3 = \dot{\Phi}_i a_i^3 \,. \tag{8.2.9}$$

We now use the freedom to rescale the scale factor to set

$$a(t_{sr}) = a_{sr} = 1 \,, \tag{8.2.10}$$

this normalization is particularly convenient to establish when a particular mode crosses the Hubble radius during slow roll, an important assessment in the analysis below.

In terms of these definitions and eqn. (8.2.9), we have that during the fast roll stage

$$\dot{\Phi}(t) = \frac{\dot{\Phi}_{sr}}{a^3(t)} \,. \tag{8.2.11}$$

Introducing

$$H_{sr}^2 \equiv \frac{V_{sr}}{3M_{Pl}^2} \,, \tag{8.2.12}$$

Friedmann's equation becomes

$$\frac{\dot{a}(t)}{a(t)} = H_{sr}\left[1 + \frac{\epsilon_V}{3\,a^6(t)}\right]^{1/2} \,. \tag{8.2.13}$$

This equation for the scale factor can be readily integrated to yield the solution

$$a(t) = \left[\left(\frac{\epsilon_V}{3}\right)^{1/2} \sinh[\theta(t)]\right]^{1/3} \quad ; \quad \theta(t) = \theta_0 + 3H_{sr}t \tag{8.2.14}$$

where θ_0 is an integration constant chosen to be

$$e^{-\theta_0} = \sqrt{\frac{\epsilon_V}{12}}\,, \tag{8.2.15}$$

so that at long time $a(t) = e^{H_{sr}t}$. The slow roll stage begins when $a(t_{sr}) = 1$ which corresponds to the value of $\theta_{sr} = \theta(t_{sr})$ given by

$$e^{-\theta_{sr}} = f\big(\frac{\epsilon_V}{3}\big) \tag{8.2.16}$$

where to simplify notation later we defined

$$f(s) = \frac{\sqrt{s}}{1+\sqrt{1+s}}\,. \tag{8.2.17}$$

Introducing the dimensionless ratio of kinetic to potential contributions at the initial time t_i

$$\frac{\dot\Phi_i^2}{2V_{sr}} = \kappa\,, \tag{8.2.18}$$

and *assuming* that the potential does not vary very much between the initial time and the onset of slow roll (this is quantified below), it follows from (8.2.9) that

$$a_i^6 = \frac{\dot\Phi_{sr}^2}{2V_{sr}\kappa} = \frac{\epsilon_V}{3\kappa} \tag{8.2.19}$$

where we have used (8.2.8). Combining this result with (8.2.14), we find that at the initial time $\theta_i = \theta(t_i)$ is given by

$$e^{-\theta_i} = f(\kappa)\,. \tag{8.2.20}$$

Let us introduce

$$\varepsilon(t) = -\frac{\dot H}{H^2} = \frac{\dot\Phi^2}{2M_{Pl}^2H^2} = \frac{\epsilon_V}{a^6(t)+\frac{\epsilon_V}{3}} \tag{8.2.21}$$

where we have used the results (8.2.7,8.2.8,8.2.9) from which it is clear that for $\epsilon_V \ll 1$ the slow roll stage begins at $a = 1$ when $\varepsilon = \epsilon_V + \mathcal{O}(\epsilon_V^2)$. With $a(t)$ given by (8.2.14), it follows that

$$\varepsilon(t) = \frac{3}{\cosh^2[\theta(t)]}\,, \tag{8.2.22}$$

therefore $0 \le \varepsilon \le 3$, and

$$H(t) = \frac{H_{sr}}{\tanh[\theta(t)]}\,. \tag{8.2.23}$$

Before we continue with the analysis, it is important to establish the relative variation of the potential between the initial time and the onset of slow roll, assuming that the potential is monotonic and does not feature "bumps", this is given by

$$\frac{\Delta V}{V_{sr}} = \Big(\frac{V'_{sr}}{V_{sr}}\Big)\,\Delta\Phi \tag{8.2.24}$$

where

$$\Delta\Phi = \int_{t_i}^{t_{sr}} \dot\Phi(t)dt = \dot\Phi_{sr}\int_{t_i}^{t_{sr}} \frac{dt}{a^3(t)} = \frac{\dot\Phi_{sr}}{3H_{sr}}\Big(\frac{3}{\epsilon_V}\Big)^{1/2}\int_{\theta_i}^{\theta_{sr}} \frac{d\theta}{\sinh[\theta]} \tag{8.2.25}$$

with the result

$$\Delta\Phi = \frac{\dot\Phi_{sr}}{3H_{sr}}\Big(\frac{3}{\epsilon_V}\Big)^{1/2}\left\{\ln\left[\frac{1+f(\kappa)}{1-f(\kappa)}\right] - \ln\left[\frac{1+f(\epsilon_V/3)}{1-f(\epsilon_V/3)}\right]\right\}. \tag{8.2.26}$$

Using (8.2.4, 8.2.8) and (8.2.17) we find

$$\Big|\frac{\Delta V}{V_{sr}}\Big| = 2\sqrt{\frac{\epsilon_V}{3}}\ln\left[\frac{1+f(\kappa)}{1-f(\kappa)}\right] + \mathcal{O}(\epsilon_V)\,. \tag{8.2.27}$$

For large κ it follows from (8.2.17) that $f(\kappa) \simeq 1 - 1/\sqrt{\kappa}$, hence the logarithm of the term in brackets varies between $1-3$ for $1 \leq \kappa \leq 100$, therefore the relative change of the potential during the fast roll stage is $\Delta V/V \simeq \sqrt{\epsilon_V}$ for $1 \leq \kappa \lesssim 100$. This result will be used in the next section below to study a systematic expansion in ϵ_V to match with the slow roll results.

The acceleration equation written in terms of $\varepsilon(t)$ is given by

$$\frac{\ddot a}{a} = H^2(t)(1-\varepsilon(t))\,, \tag{8.2.28}$$

so that the inflationary stage begins when $\varepsilon(t) = 1$. At the initial time

$$\varepsilon(t_i) = \frac{3\kappa}{1+\kappa} \tag{8.2.29}$$

hence, for $\kappa > 1/2$ the early stage of expansion is deccelerated and inflation begins when $\varepsilon(t_{inf}) = 1$.

It proves convenient to introduce the variable

$$x(t) = e^{-\theta(t)/3} = \left[\frac{\epsilon_V}{12}\right]^{1/6} e^{-H_{sr}t}\,, \tag{8.2.30}$$

with

$$x_i \equiv x(t_i) = [f(\kappa)]^{1/3} \quad ; \quad x_{sr} \equiv x(t_{sr}) = [f(\epsilon_V/3)]^{1/3} \tag{8.2.31}$$

where $f(s)$ is given by eqn. (8.2.17), and write a, H, ε in terms of this variable leading to

$$a(x) = \left[\frac{\epsilon_V}{12}\right]^{1/6} \frac{\left[1-x^6\right]^{1/3}}{x}\,, \tag{8.2.32}$$

$$H(x) = H_{sr}\frac{\left[1+x^6\right]}{\left[1-x^6\right]}\,, \tag{8.2.33}$$

$$\varepsilon(x) = \frac{12\,x^6}{\left[1+x^6\right]^2}\,. \tag{8.2.34}$$

The number of e-folds between the initial time t_i and a given time t is given by

$$N_e(t;t_i) = \int_{t_i}^{t} H(t')\,dt' = \frac{1}{3}\,\ln\left[\sqrt{\kappa}\,\frac{(1-x^6(t))}{2x^3(t)}\right], \tag{8.2.35}$$

with a total number of e-folds between the beginning of the fast roll stage at $t = t_i$ and the onset of slow roll at t_{sr} given by

$$N_e(t_i;t_{sr}) = \frac{1}{6}\,\ln\left[\frac{3\kappa}{\epsilon_V}\right]. \tag{8.2.36}$$

Fig. (36) shows ε as a function of N_e for $\kappa = 10; 100, \epsilon_V = 0.008$, inflation begins at $N_e \simeq 0.5 - 0.8$ and slow roll begins at $N_e \simeq 1.37 - 1.75$. We find that this is the typical behavior for $1 \leq \kappa \leq 100$, namely for a wide range of fast roll initial conditions, the inflationary stage begins fairly soon $N_{e,inf} \lesssim 1$ and the fast roll stage lasts $\lesssim 1.7$ e-folds.

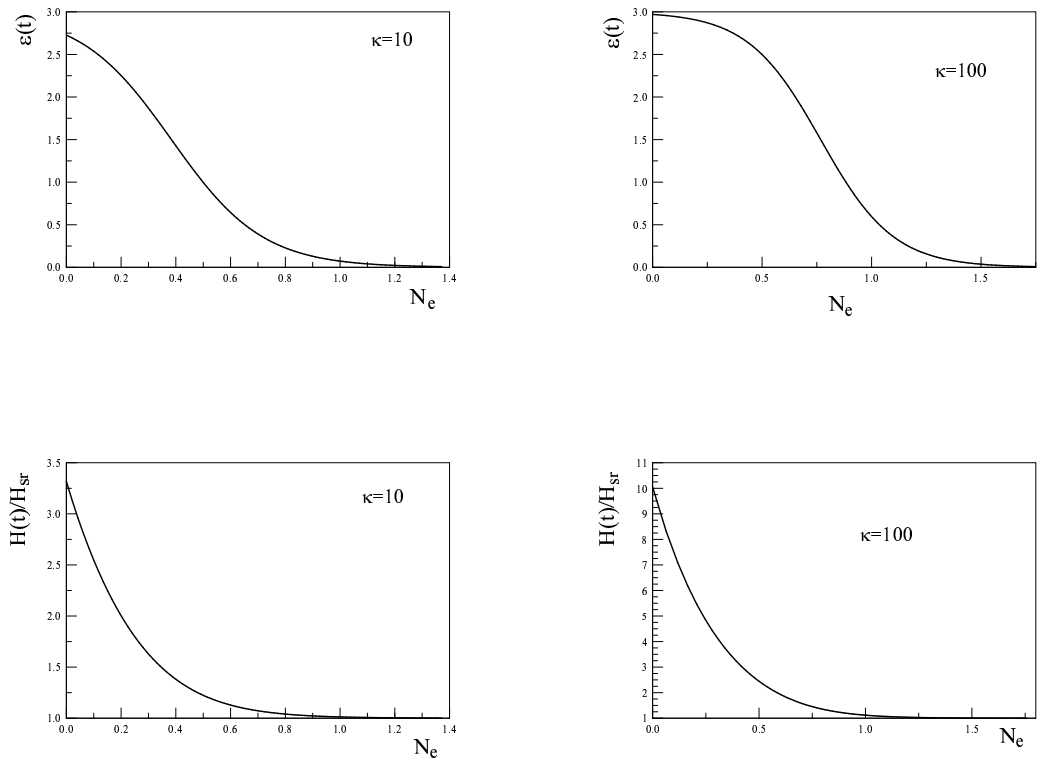


Figure 36: $\varepsilon(t)$ and $H(t)/H_{sr}$ as a function of the number of e-folds from the beginning of fast roll, for $\kappa = 10; 100$ for $\epsilon_V = 0.008$. Inflation starts at $N_e \simeq 0.5$, slow roll starts at $N_e \lesssim 1.75$.

8.3 MATCHING TO SLOW ROLL:

At the end of the fast roll stage the value of $\dot{\Phi}^2/2V_{sr} \simeq \mathcal{O}(\epsilon_V)$, which is of the same order as the slow-roll solution of the equations of motion, and becomes *smaller* than the slow roll solution for $t > t_{sr}$ for which $a(t) > 1$. Therefore we must ensure a smooth matching to the slow roll stage. This is accomplished by recognizing that with the fast roll initial conditions there emerges a hierarchy of time scales as well as amplitudes for $\dot{\Phi}$: during the fast roll

stage the $\dot{\Phi}$ features a large amplitude $\propto \sqrt{\kappa} \gg 1$ and varies fast, while in the slow roll stage the amplitude is $\propto \sqrt{\epsilon_V} \ll 1$ and varies slowly. Furthermore, in the previous section we have taken the potential to be (nearly) constant and recognized that the relative variation during the fast roll stage is $\Delta V/V_{sr} \propto \sqrt{\epsilon_V}$.

In this section we treat the variation of the potential along with the slow roll corrections in a consistent perturbative formulation.

Therefore we write

$$\Phi(t) = \Phi^f(t) + \Phi^s(t) \quad ; \quad \Phi^f \equiv \Phi^{(0)}, \Phi^s \equiv \Phi^{(1)} + \Phi^{(2)} + \cdots \tag{8.3.1}$$

where *formally* $\dot{\Phi}^{(0)} \propto (\sqrt{\epsilon_V})^0$; $\dot{\Phi}^{(1)} \propto (\sqrt{\epsilon_V})$; $\dot{\Phi}^{(2)} \propto (\epsilon_V) \cdots$ with $\dot{\Phi}^{(0)}(t)$ being the fast roll solution (8.2.11) which is of amplitude $\sqrt{\kappa}$ during most of the fast roll stage. Furthermore, during the fast roll stage we assumed that the potential is nearly constant and equal to the potential during the slow roll stage, namely $V(\Phi) \simeq V_{sr}$. We now relax this assumption by writing $\Phi = \Phi_{sr} + (\Phi - \Phi_{sr})$ in the argument of the potential $V(\Phi)$ as in eqns. (8.2.24-8.2.27)

$$V(\Phi) = V_{sr} + \Delta V(t) \quad ; \quad \Delta V(t) = V'_{sr}\Big(\Delta\Phi(t)\Big) + \frac{1}{2}V''_{sr}\Big(\Delta\Phi(t)\Big)^2 + \cdots . \tag{8.3.2}$$

where

$$\begin{aligned} \Delta\Phi(t) &= \int_\infty^t \Big(\dot{\Phi}(t') - \dot{\Phi}_{sr}(t')\Big)dt' = \Delta\Phi^{(0)}(t) + \Delta\Phi^{(1)}(t) + \cdots \\ \Delta\Phi^{(0)}(t) &= \int_\infty^t \dot{\Phi}^{(0)}(t')dt' \quad ; \quad \Delta\Phi^{(1)}(t) = \int_\infty^t \Big(\dot{\Phi}^{(1)}(t') - \dot{\Phi}_{sr}(t')\Big)dt' . \end{aligned} \tag{8.3.3}$$

In writing this expression, we have assumed (self-consistently, see below) that at long time $\dot{\Phi}(t) \to \dot{\Phi}_{sr}(t)$, hence adjusting an integration constant asymptotically at long time $\Phi(t) \to \Phi_{sr}(t)$ thereby extending the lower limit of the integral to $t \to \infty$. This assumption will be justified *a posteriori* from the solution.

Formally V'_{sr} is of $\mathcal{O}(\sqrt{\epsilon_V})$, $V''_{sr} \propto \eta_V$ is of $\mathcal{O}(\epsilon_V)$, etc. Therefore

$$\Delta V = \Delta V^{(1)} + \Delta V^{(2)} + \cdots , \tag{8.3.4}$$

with

$$\Delta V^{(1)} = V'_{sr}\ \Delta\Phi^{(0)}(t) \tag{8.3.5}$$

$$\Delta V^{(2)} = V'_{sr}\ \Delta\Phi^{(1)}(t) + \frac{1}{2}V''_{sr}\Big[\Delta\Phi^{(0)}(t)\Big]^2 \tag{8.3.6}$$

$$\vdots = \vdots \tag{8.3.7}$$

Similarly we write

$$H(t) = \frac{1}{3M_{Pl}^2}\Big[\frac{1}{2}\Big(\dot\Phi^{(0)} + \dot\Phi^{(1)} + \cdots\Big)^2 + V_{sr} + \Delta V\Big]^{1/2} \equiv H^{(0)} + H^{(1)} + H^{(2)} + \cdots\,, \tag{8.3.8}$$

where $H^{(0)}$ is the fast roll solution (8.2.13) with (8.2.12,8.2.14) and

$$H^{(1)} = \frac{H_{sr}^2}{H^{(0)}}\Big[\frac{\dot\Phi^{(0)}\dot\Phi^{(1)}}{2V_{sr}} + \frac{\Delta V^{(1)}}{V_{sr}}\Big]\ \ ;\ \ H^{(2)} = \frac{H_{sr}^2}{H^{(0)}}\Big[\frac{\big(\dot\Phi^{(1)}\big)^2}{2V_{sr}} + \frac{\Delta V^{(2)}}{V_{sr}}\Big]\ \ ;\ \ \text{etc}\,. \tag{8.3.9}$$

With the fast roll solution (8.2.11,8.2.14) we find that

$$\Delta\Phi^{(0)}(t) = \frac{\dot\Phi_{sr}}{3H_{sr}}\left(\frac{3}{\epsilon_V}\right)^{1/2}\ln\Big[\frac{1-x^3}{1+x^3}\Big]\,, \tag{8.3.10}$$

where $x(t)$ is given by eqn. (8.2.30). From this we obtain

$$\frac{\Delta V^{(1)}}{V_{sr}} = 2\sqrt{\frac{\epsilon_V}{3}}\ln\Big[\frac{1+x^3}{1-x^3}\Big]\,, \tag{8.3.11}$$

$$\frac{V''_{sr}}{2V_{sr}}\Big[\Delta\Phi^{(0)}(t)\Big]^2 = \frac{\eta_V}{3}\ln^2\Big[\frac{1+x^3}{1-x^3}\Big]\,. \tag{8.3.12}$$

In the equation of motion

$$\ddot\Phi + 3H\dot\Phi = -V'(\Phi) \tag{8.3.13}$$

the right hand side is *formally* of order $-V'(\Phi) \propto \sqrt{\epsilon_V} + \cdots$ as can be seen from the definition of the slow roll variable ϵ_V (8.2.4). This suggests an expansion in powers of $\sqrt{\epsilon_V}$ which leads to the following hierarchy of equations

$$\ddot\Phi^{(0)} + 3H^{(0)}\dot\Phi^{(0)} = 0 \tag{8.3.14}$$

$$\ddot\Phi^{(1)} + 3H^{(0)}\dot\Phi^{(1)} + 3H^{(1)}\dot\Phi^{(0)} = -V'_{sr} \tag{8.3.15}$$

$$\vdots = \vdots \tag{8.3.16}$$

Consistently with the slow roll approximation and to lowest order in slow roll we neglect $\ddot{\Phi}^{(1)}$ in (8.3.15) as it can be shown (a posteriori) that $\ddot{\Phi}^{(1)} \propto \left[3H^{(0)}\dot{\Phi}^{(1)}\right](\epsilon_V + \eta_V)$, hence higher order in the slow roll expansion.

Inserting the results (8.2.11) and (8.2.14) for the zeroth order fast roll solution, and the result (8.3.11) along with the leading order slow roll relations (8.2.8) and the slow roll result

$$\dot{\Phi}_{sr} = -\frac{V'_{sr}}{3H_{sr}} \tag{8.3.17}$$

into eqn. (8.3.15) (neglecting $\ddot{\Phi}^{(1)}$), we find

$$\frac{\dot{\Phi}^{(1)}}{\dot{\Phi}_{sr}} \equiv F[x] = \frac{T[x]}{2 - T^2[x]} \left\{ 1 - \frac{2x^3}{3(1+x^6)} \ln\left[\frac{1+x^3}{1-x^3}\right] \right\} \;\; ; \;\; T[x] = \frac{1-x^6}{1+x^6} \tag{8.3.18}$$

The function $F[x]$ features the following asymptotic behavior,

$$F \simeq \frac{1}{\sqrt{\kappa}} \left\{ 1 + \frac{1}{6} \ln\left[\frac{\kappa}{4}\right] \right\} \;\; \text{for} \;\; t \to t_i \; , \; \kappa \gg 1 \tag{8.3.19}$$

$$F \simeq 1 + \mathcal{O}(x^6) \;\; \text{for} \;\; t \geq t_{sr} \, . \tag{8.3.20}$$

Therefore for $t \to t_i$ it follows that

$$\frac{\dot{\Phi}^{(1)}}{\dot{\Phi}^{(0)}} \simeq \frac{1}{12}\sqrt{\frac{\epsilon_V}{12}} \, \frac{\ln\left[\kappa\right]}{\kappa} \tag{8.3.21}$$

$$\frac{H^{(1)}}{H^{(0)}} \simeq \sqrt{\frac{\epsilon_V}{12}} \, \frac{\ln\left[\kappa\right]}{\kappa} \, . \tag{8.3.22}$$

With the results obtained above, it is straightforward to confirm that the second order correction is indeed of $\mathcal{O}(\epsilon_V, \eta_V)$ and further suppressed by a power of κ up to logarithmic terms in κ.

Therefore up to first order in slow roll

$$\dot{\Phi} = \dot{\Phi}_{sr}\left[\frac{1}{a^3} + F[x]\right] , \tag{8.3.23}$$

$$H^2 = H^2_{sr}\left\{ 1 + \frac{\epsilon_V}{3}\left[\frac{1}{a^3} + F[x]\right]^2 \right\} . \tag{8.3.24}$$

The second order contribution $\Delta V^{(2)}/V_{sr}$ can be found by carrying out the integral in the second term in (8.3.3), this is achieved more efficiently by passing to the variable x and

expanding the function x in a series in x^3 and integrating term by term. The result reveals that this correction is $\mathcal{O}(\epsilon_V, \eta_V)$ and suppressed by a power of κ as $t \to t_i$ and is also subleading for $t \geq t_{sr}$.

The result (8.3.18) clearly shows that for $t \gtrsim t_{sr}$

$$\dot{\Phi}^{(1)} - \dot{\Phi}_{sr} = \dot{\Phi}_{sr}\Big[F[x] - 1\Big] \simeq \epsilon_V \, e^{-6H_{sr}t} \tag{8.3.25}$$

(see eqn. (8.2.30)) so that adjusting the integration constant in eqn. (8.3.25) so that $\Phi^{(1)} \to \Phi_{sr}$ as $t \to \infty$ justifies the assumption that asymptotically $\Phi - \Phi_{sr} \to 0$ as $t \to \infty$ thus validating the expressions (8.3.3).

For $t \to t_i$ these expressions reduce to the fast roll results of the previous section; however, for $t > t_{sr}$ ($x^3 \lesssim (\epsilon_V/12)^{1/2}$) we find

$$\dot{\Phi} \;\to\; \dot{\Phi}_{sr} + \mathcal{O}\Big(\epsilon_V \, e^{-6H_{sr}t}\Big) \tag{8.3.26}$$

$$H^2 \;\to\; H_{sr}^2\Big[1 + \frac{\epsilon_V}{3}\Big] \, . \tag{8.3.27}$$

The correction to the scale factor is obtained by proposing a solution of the form

$$a(t) = a^{(0)}(t) \; a^{(s)}(t) \tag{8.3.28}$$

where

$$\frac{\dot{a}^{(0)}}{a^{(0)}} = H^{(0)} \; , \; \frac{\dot{a}^{(s)}}{a^{(s)}} = H^{(1)} + H^{(2)} + \cdots \tag{8.3.29}$$

It is straightforward to find that asymptotically for $t \gg t_{sr}$

$$a^{(0)}(t) = e^{H_{sr}t} \; , \; a^{(s)}(t) \to \Big[a^{(0)}(t)\Big]^{\epsilon_V/6} \tag{8.3.30}$$

where a detailed calculation shows that the terms proportional to ΔV are subleading for $t \gg t_{sr}$. Therefore the improved fast roll solution (8.3.24) yields the *correct* long time behavior of the scale factor to leading order in slow roll; namely, for $t > t_{sr}$, the dynamics enters a near de Sitter stage

$$\ln[a(t)] \to H_{sr}\Big[1 + \frac{\epsilon_V}{6}\Big] \; t \, . \tag{8.3.31}$$

It is now clear from the solution (8.3.23) that for $t \ll t_{sr}$ the fast roll, zeroth order solution ($\propto 1/a^3$) dominates but at $t \simeq t_{sr}$ ($a(t_{sr}) = 1$), $F[x_{sr}] \simeq 1$ and $\dot{\Phi} \approx 2\dot{\Phi}_{sr}$. Therefore,

at t_{sr}, the solution is of order $\sqrt{\epsilon_V}$ but off by a factor 2 from the correct solution, resulting in an error of $\mathcal{O}(\epsilon_V)$. In order to match to the correct slow roll solution the evolution must be continued past t_{sr} to a time t_m at which the first order correction dominates. This "matching time", t_m, is determined by the error incurred in keeping the zeroth order term in the full solution. For example, requiring that the error be $\simeq \epsilon_V \sqrt{\epsilon_V}$ fixes t_m so that

$$a^3(t_m) \simeq \frac{1}{\sqrt{\epsilon_V}} \Rightarrow x_m^3 \simeq \left(\frac{\epsilon_V^2}{12}\right)^{1/2} \tag{8.3.32}$$

hence, at the "matching time", we find that

$$\dot{\Phi}(t_m) = \dot{\Phi}_{sr}\left[1 + \mathcal{O}(\sqrt{\epsilon_V})\right]. \tag{8.3.33}$$

The number of e-folds between the time t_{sr}, at which $\dot{\Phi}^{(0)} \simeq \dot{\Phi}_{sr}$, and the matching time t_m is

$$N_e(t_{sr}; t_m) = -\frac{1}{6}\ln[\epsilon_V] \simeq 0.8 \tag{8.3.34}$$

where the numerical result applies for $\epsilon_V = 0.008$. Therefore for $\kappa \lesssim 100, \epsilon_V = 0.008$, the total number of e-folds between the initial and the matching time is $N_e \simeq 2.5$.

With the improved solution (8.3.23), it follows that the variable $\varepsilon(t)$ defined by eqn. (8.2.21) is given by

$$\varepsilon(t) = \frac{\epsilon_V\left[\frac{1}{a^3} + F[x]\right]}{1 + \frac{\epsilon_V}{3}\left[\frac{1}{a^3} + F[x]\right]}. \tag{8.3.35}$$

This quantity is a better indicator of the transition to slow roll, it features the following limits

$$\begin{aligned} \varepsilon(t) &\simeq 3 \quad \text{for} \quad t \simeq t_i, \kappa \gg 1 \\ \varepsilon(t) &\simeq 2\epsilon_V \quad \text{for} \quad t \simeq t_{sr} \\ \varepsilon(t) &\simeq \epsilon_V \quad \text{for} \quad t > t_{sr}\,. \end{aligned} \tag{8.3.36}$$

with corrections of $\mathcal{O}(\epsilon_V^{3/2})$ at the matching time t_m.

Conformal time $\eta(t)$ defined to vanish as $t \to \infty$ is given by

$$\begin{aligned} \eta(t) &= \int_\infty^t \frac{dt'}{a(t')} = \int_\infty^{a(t)} \frac{da}{a^2\, H(a)} \\ &= -\frac{1}{a(t)H(t)} + \int_\infty^t \varepsilon(t') \frac{dt'}{a(t')} \end{aligned} \tag{8.3.37}$$

where we integrated by parts and used the definition of ε given by eqn. (8.2.21). Adding and subtracting ϵ_V we find

$$\eta(t) = -\frac{1}{a(t)H(t)(1-\epsilon_V)} + \frac{\epsilon_V}{\big(1-\epsilon_V\big)} \int_\infty^t \Big[\frac{\varepsilon(t')}{\varepsilon_V} - 1\Big] \frac{dt'}{a(t')}\,, \tag{8.3.38}$$

The argument of the integrand in the second term in (8.3.38) vanishes to leading order in ϵ_V, η_V in the slow roll phase (when $t > t_{sr}$). Therefore, during slow roll, $\eta = -1/aH(1-\epsilon_V)$.

Discussion:

The study in this section describes a systematic procedure to obtain a solution that is valid during the fast roll stage and that matches smoothly to the slow roll stage to any desired order in ϵ_V, η_V *independently of the potential* while under the assumption that the inflationary potential is monotonic and can be described by a derivative expansion characterized by slow roll parameters. The leading order solution is the fast roll solution (obtained in the previous section) and the above analysis shows that continuing this solution for time larger than t_{sr} incurs errors of order ϵ_V in the variable ε: at $t = t_{sr}$ the zeroth-order and the improved solution differ by ϵ_V which in turn leads to corrections $\leq \epsilon_V^2$ in the conformal time η.

This analysis shows that the leading order corrections to the inflationary power spectra from a fast roll stage can be obtained by keeping only the fast-roll solution and integrating up to $t \simeq t_{sr}$, at which point it matches to slow roll. Clearly keeping only the zeroth-order solution rather than the improved solution incurs errors of $\mathcal{O}(\epsilon_V)$, which can (with numerical effort) be systematically improved upon by considering the corrections and improvements described in this section.

Having quantified the error incurred in keeping only the fast roll solution, we now proceed to obtain the corrections to the power spectra of scalar and tensor perturbations *to leading order in the expansion in slow roll parameters*, namely keeping only the fast roll solution.

8.4 FAST ROLL CORRECTIONS TO POWER SPECTRA:

The analysis above clearly indicates that for a wide range of initial conditions dominated by the kinetic term of the inflaton potential, a fast roll stage merges with the slow roll stage within $2-3$ e-folds. Having quantified above the error incurred in keeping only the fast roll solution, we now proceed to obtain the corrections to the power spectra of scalar and tensor perturbations *to leading order in the expansion in slow roll parameters*, namely keeping only the fast roll solution. The results will yield the main corrections to the power spectra from the fast roll stage, with potential corrections of $\mathcal{O}(\epsilon_V)$ from the matching of scales. If the main features of the results obtained in leading order are supported observationally, this would justify a more thorough study that includes these corrections by implementing the systematic approach described in the previous section. Such a program would necessarily imply a larger numerical effort and would be justified if observational data suggest the presence of the main effects.

The observational constraint of nearly scale invariance suggests that wavelengths corresponding to observable quantities today crossed the Hubble radius during the slow roll era of inflation. Therefore our goal is to analyze the impact of the pre-slow roll dynamics upon perturbations with physical wavelengths that crossed the Hubble radius *after* the beginning of slow roll. As discussed in refs.[275, 276] and more recently in ref.[311] the fast-roll stage prior to slow roll modifies the initial conditions on the mode functions from the usual Bunch-Davies case. The rapid dynamical evolution of the inflaton during the fast roll stage induces a correction to the potential in the equations of motion for the mode functions of curvature and tensor perturbations, which we now analyze in detail.

The gauge invariant curvature perturbation of the comoving hypersurfaces is given in terms of the Newtonian potential ($\psi(\vec{x},t)$) and the inflaton fluctuation ($\delta\phi(\vec{x},t)$) by[241, 245, 242, 243, 244]

$$\mathcal{R} = -\psi - \frac{H}{\dot{\Phi}}\,\delta\phi\ . \tag{8.4.1}$$

where $\dot{\Phi}$ stands for the derivative of the inflaton field Φ with respect to the cosmic time t.

It is convenient to introduce the gauge invariant potential [241, 245, 242, 243, 244],

$$u(\vec{x},t) = -z\,\mathcal{R}(\vec{x},t)\ , \tag{8.4.2}$$

where

$$z = a(t)\,\frac{\dot{\Phi}}{H}\ . \tag{8.4.3}$$

The gauge invariant field $u(\vec{x},t)$ is quantized by expanding in terms of conformal time mode functions and creation and annihilation operators as follows[241, 245, 242, 243, 244]

$$u(\vec{x},\eta) = \frac{1}{\sqrt{V}}\sum_{\vec{k}}\left[\alpha_{\mathcal{R}}(k)\,S_{\mathcal{R}}(k;\eta)\,e^{i\vec{k}\cdot\vec{x}} + \alpha^{\dagger}_{\mathcal{R}}(k)\,S^{*}_{\mathcal{R}}(k;\eta)\,e^{-i\vec{k}\cdot\vec{x}}\right]. \tag{8.4.4}$$

The operators $\alpha_{\mathcal{R}}(k), \alpha^{\dagger}_{\mathcal{R}}(k)$ obey canonical commutation relations and the mode functions are solutions of the equation

$$\left[\frac{d^2}{d\eta^2} + k^2 - \frac{z''}{z}\right] S_{\mathcal{R}}(k;\eta) = 0\,. \tag{8.4.5}$$

Tensor perturbations (gravitational waves) correspond to minimally coupled massless fields with two physical transverse polarizations, the quantum fields are written as [241, 245, 242, 243, 244]

$$h_{ij}(\vec{x},\eta) = \frac{2}{C(\eta)\,M_{Pl}}\sum_{\vec{k}}\sum_{\lambda=\times,+}\epsilon_{ij}(\lambda,\vec{k})\left[\alpha_{\lambda,\vec{k}}\ S_T(k;\eta)\,e^{i\vec{k}\cdot\vec{x}} + \alpha^{\dagger}_{\lambda,\vec{k}}\ S^{*}_T(k;\eta)\,e^{-i\vec{k}\cdot\vec{x}}\right]\ , \tag{8.4.6}$$

where λ labels the two standard transverse and traceless polarizations $\times$ and $+$. The operators $\alpha_{\lambda,\vec{k}},\ \alpha^{\dagger}_{\lambda,\vec{k}}$ obey the usual canonical commutation relations, and $\epsilon_{ij}(\lambda,\vec{k})$ are the two independent traceless-transverse tensors constructed from the two independent polarization vectors transverse to $\hat{\mathbf{k}}$, chosen to be real and normalized such that $\epsilon^{i}_{j}(\lambda,\vec{k})\ \epsilon^{j}_{k}(\lambda',\vec{k}) = \delta^{i}_{k}\ \delta_{\lambda,\lambda'}$.

The mode functions $S_T(k;\eta)$ obey the differential equation of a massless minimally coupled scalar field, namely

$$\left[\frac{d^2}{d\eta^2} + k^2 - \frac{C''(\eta)}{C(\eta)}\right] S_T(k;\eta) = 0\,. \tag{8.4.7}$$

In both these cases the mode functions obey an equation of the form,

$$\left[\frac{d^2}{d\eta^2} + k^2 - W_{\alpha}(\eta)\right] S_{\alpha}(k;\eta) = 0\ \ ;\ \ \alpha = R, T\,. \tag{8.4.8}$$

This is a Schrödinger equation with η playing the role of coordinate, k^2 the energy and $W(\eta)$ a potential that depends on the coordinate η. In the cases under consideration

$$W_\alpha(\eta) = \begin{cases} z''/z & \text{for curvature perturbations} \\ C''/C & \text{for tensor perturbations} \end{cases} . \tag{8.4.9}$$

During slow roll inflation the potential $W_\alpha(\eta)$ becomes

$$W_\alpha(\eta) = \frac{\nu_\alpha^2 - \frac{1}{4}}{\eta^2} , \tag{8.4.10}$$

where to leading order in slow roll parameters

$$\nu_\alpha = \frac{3}{2} + \begin{cases} 3\epsilon_V - \eta_V & \text{for curvature perturbations} \\ \epsilon_V & \text{for tensor perturbations} \end{cases} . \tag{8.4.11}$$

The full dynamical evolution of the inflaton during the fast roll stage leads to a modification of the mode equations (8.4.8) over terms of a potential $V_\alpha(\eta)$ that is localized in η in a narrow range prior to the slow roll phase[275, 276, 277]. Specifically, in the mode equations (8.4.8), $W(\eta)$ is modified as

$$W_\alpha(\eta) = V_\alpha(\eta) + \frac{\nu_\alpha^2 - 1/4}{\eta^2} \quad ; \quad V_\alpha(\eta) = \begin{cases} \neq 0 \text{ for } & \eta_i < \eta < \eta_{sr} \\ 0 \text{ for } & \eta_{sr} < \eta \, , \end{cases} \tag{8.4.12}$$

where ν_α is given by (8.4.11) for curvature and tensor perturbations.

For curvature perturbations we find

$$W_R(\eta) = \frac{z''}{z} = \frac{a^2}{z}[\ddot{z} + H\dot{z}] = 2a^2H^2\left[1 - \frac{7}{2}\varepsilon + \varepsilon^2 + (3-\varepsilon)\left[2\sqrt{\varepsilon\epsilon_V} - \frac{\eta_V}{2}\right]\right] \tag{8.4.13}$$

Therefore, to leading order in slow roll, the potential for curvature perturbations is given by

$$V_R(\eta) = W_R(\eta) - \frac{2}{\eta^2}\left[1 + \frac{9}{2}\epsilon_V - \frac{3}{2}\eta_V\right] . \tag{8.4.14}$$

For tensor perturbations

$$W_T(\eta) = \frac{C''}{C} = a[\ddot{a} + H\dot{a}] = 2a^2H^2\left[1 - \frac{\varepsilon}{2}\right] , \tag{8.4.15}$$

and the potential for tensor perturbations is given by

$$V_T(\eta) = W_T(\eta) - \frac{2}{\eta^2}\Big[1 + \frac{3\epsilon_V}{2}\Big]\,, \tag{8.4.16}$$

The potentials as a function of η are found parametrically in terms of the dimensionless variable x, given by (8.2.30), by writing a, H, ε, η all as functions of x.

From the result of the previous sections, it is clear that as $t \to t_{sr}$, $\varepsilon \to \epsilon_V$ and $a^2H^2 \to 1/\eta^2$, higher order corrections in ϵ_V arise from the matching to the slow roll stage thereby yielding higher order corrections in ϵ_V, η_V to the potentials $V_{R,T}$.

Therefore we focus on obtaining the leading order effects from the fast-roll stage by considering solely the fast-roll solution given by eqns.(8.2.32-8.2.34) along with replacing the fast roll solution (8.2.21) into the expression for η (8.3.38). This yields

$$\begin{aligned}\eta(x) &= -\frac{1}{H_{sr}\left(1-\epsilon_V\right)}\left(\frac{12}{\epsilon_V}\right)^{1/6}\Bigg\{\frac{x(1-x^6)^{2/3}}{(1+x^6)} \\ &+ \epsilon_V\int_{x_{sr}}^{x}\frac{dy}{[1-y^6]^{1/3}}\left[\frac{12}{\epsilon_V}\frac{y^6}{(1+y^6)^2}-1\right]\Bigg\}\,.\end{aligned} \tag{8.4.17}$$

where the lower limit in the integral ensures the matching to the slow roll result at t_{sr}. The potentials are now obtained to leading order by replacing the expressions (8.2.32-8.2.34,8.4.17) into (8.4.13,8.4.14) and in (8.4.15,8.4.16). As discussed in the previous sections, considering the lowest order solutions captures the full fast roll stage and yields an error $\simeq \mathcal{O}(\epsilon_V)$ for $t > t_{sr}$ during the slow roll stage.

The potentials $V_R(\eta)$; $V_T(\eta)$ are shown in figs. (37,38) for $\kappa = 10; 100$ for $\epsilon_V = 0.008$; $\eta_V = -0.010$.

These potentials are qualitatively similar to those found for a specific choice of the inflaton potential and initial conditions in the second reference in[275]. Two important aspects explain most of the quantitative discrepancy between our results and those of this reference: i) the potential scales $\propto a^2$, therefore by normalizing the scale factor to unity at the beginning of slow roll, our potential features an overall scale with respect to that in ref.[275], ii) η scales $\propto 1/a$ therefore there is also an overall scaling in the definition of conformal time, our normalization is more convenient to analyze the transition to slow roll

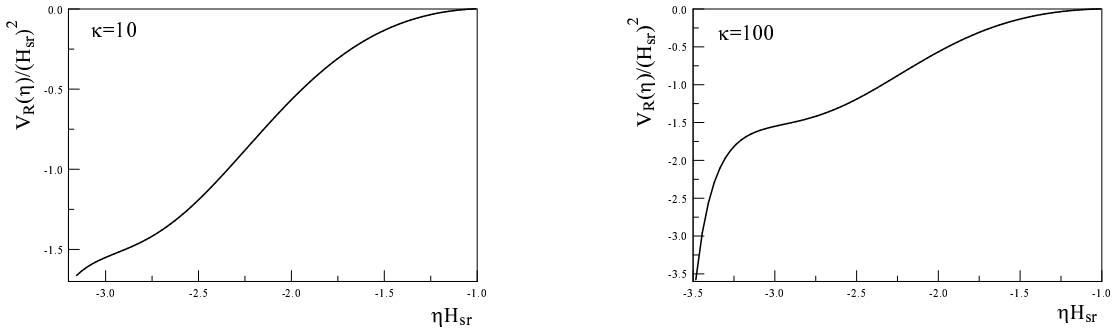


Figure 37: Potentials for curvature perturbations $V_R(\eta)$ as a function of η from the beginning of fast roll, for $\kappa = 10; 100$; $\epsilon_V = 0.008$; $\eta_V = -0.010$.

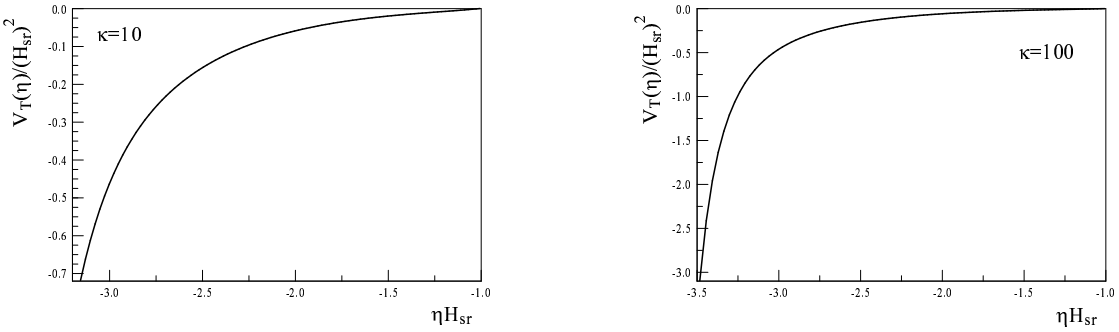


Figure 38: Potentials for tensor perturbations $V_T(\eta)$ as a function of η from the beginning of fast roll, for $\kappa = 10; 100$; $\epsilon_V = 0.008$; $\eta_V = -0.010$.

and assess when wavevector cross the Hubble radius during slow roll inflation. Furthermore the particular choice of the potential in this reference also modifies the values of $\epsilon_V, \eta_V; \kappa$. Accounting for the different normalizations (scale of a affecting the definition of the potential

and η), the similarity of the potentials is both reassuring and expected because the potential is determined by the fast roll stage which is dominated by the fast evolution of the inflaton field and is rather insensitive to the potential as long as the potential is sufficiently flat to be consistent with slow roll. Thus our results are robust and to leading order in slow roll variables only depend on κ, ϵ_V and η_V *regardless of the specific form of the inflationary potential*.

During the slow roll stage the solution of the mode equations (8.4.8) with $V_\alpha(\eta) = 0$, namely with $W_\alpha(\eta)$ given by (8.4.10), are

$$S_\alpha(k;\eta) = A_{k,\alpha}\, g_{\nu_\alpha}(k;\eta) + B_{k,\alpha}\, g^*_{\nu_\alpha}(k;\eta) \quad ; \quad \alpha = T, R\,. \tag{8.4.18}$$

where, up to an overall phase,

$$g_\nu(k,\eta) = \sqrt{\frac{-\pi\eta}{4}}\, H^{(1)}_\nu(-k\eta) \tag{8.4.19}$$

are the solutions with Bunch-Davies initial conditions, where ν_α is given by (8.4.11) for curvature and tensor perturbations.

The power spectra for curvature ($\mathcal{R}$) and tensor (gravitational waves) (T) perturbations are respectively

$$\mathcal{P}_\mathcal{R}(k) = \frac{k^3}{2\pi^2}\left|\frac{S_\mathcal{R}(k;\eta)}{z(\eta)}\right|^2 \quad ; \quad \mathcal{P}_T(k) = \frac{4\,k^3}{\pi^2\, M_{pl}^2}\left|\frac{S_T(k;\eta)}{C(\eta)}\right|^2 . \tag{8.4.20}$$

We assume that the modes of cosmological relevance today crossed the Hubble radius during the slow roll inflationary stage, therefore evaluating these power spectra a few e-folds after horizon crossing $-k\eta \ll 1$ during slow roll, it follows that the general solution (8.4.18) is given by

$$S_\alpha(k;\eta) = -i\sqrt{\frac{-\pi\eta}{4}}\,\frac{\Gamma(\nu_\alpha)}{\pi}\Big(-\frac{k\eta}{2}\Big)^{-\nu_\alpha}\big[A_{k,\alpha} - B_{k,\alpha}\big] \quad ; \quad -k\eta \ll 1\,, \tag{8.4.21}$$

therefore the power spectra become

$$\mathcal{P}_\alpha(k) = \mathcal{P}^{BD}_\alpha(k)\,\mathcal{T}_\alpha(k) \quad ; \quad \alpha = \mathcal{R}, T\,, \tag{8.4.22}$$

where $\mathcal{P}_\alpha^{BD}(k)$ are the power spectra for Bunch-Davies modes $g_\nu(k;\eta)$, namely for $A_k = 1; B_k = 0$, and

$$\mathcal{T}_\alpha(k) = \left|A_{k,\alpha} - B_{k,\alpha}\right|^2 \tag{8.4.23}$$

is a transfer function that encodes the non-Bunch-Davies initial conditions for the respective perturbations.

With $W_\alpha(\eta)$ given by (8.4.10,8.4.11) during slow roll, we find

$$z(\eta) = z_0\Big(\frac{-\eta}{-\eta_0}\Big)^{\frac{1}{2}-\nu_\mathcal{R}} \quad ; \quad z_0 = \left[\frac{a\dot\Phi}{H}\right]_{\eta_0}, \tag{8.4.24}$$

and

$$C(\eta) = C_0\Big(\frac{-\eta}{-\eta_0}\Big)^{\frac{1}{2}-\nu_T} \quad ; \quad C_0 = \left[\frac{1}{-\eta\, H}\right]_{\eta_0}, \tag{8.4.25}$$

where C_0 given in eqn. (8.4.25) is to leading order in slow roll and η_0 is an arbitrary scale. Therefore, to leading order in slow roll, we find that

$$\mathcal{P}_\mathcal{R}(k) = \frac{1}{4\pi^2\, z_0^2\, \eta_0^2}\Big(-k\eta_0\Big)^{n_s-1}\mathcal{T}_\mathcal{R}(k) \quad ; \quad n_s - 1 = -6\epsilon_V + 2\eta_V\,. \tag{8.4.26}$$

Using the value z_0 in eqn.(8.4.24), along with the slow roll relation $\dot\Phi^2/H^2 = 2M_{Pl}^2\epsilon_V$ and defining $-\eta_0 \equiv 1/k_0$ as a "pivot" scale, we finally find to leading order in slow roll

$$\mathcal{P}_\mathcal{R}(k) = \frac{H^2}{8\pi^2\, M_{Pl}^2\epsilon_V}\Big|_{\eta_0=-1/k_0}\Big(\frac{k}{k_0}\Big)^{n_s-1}\mathcal{T}_\mathcal{R}(k)\,. \tag{8.4.27}$$

Therefore, choosing k as the "pivot" scale k_0, gives us

$$\mathcal{P}_\mathcal{R}(k_0) = \frac{H^2}{8\pi^2\, M_{Pl}^2\epsilon_V}\Big|_{-\eta_0=1/k_0}\mathcal{T}_\mathcal{R}(k_0)\,. \tag{8.4.28}$$

Carrying out similar steps for tensor perturbations we find

$$\mathcal{P}_T(k) = \frac{2}{\pi^2\, M_{Pl}^2 C_0^2\, \eta_0^2}\Big(-k\eta_0\Big)^{n_T}\mathcal{T}_T(k) \quad ; \quad n_T = -2\epsilon_V\,. \tag{8.4.29}$$

Using $C_0^2\eta_0^2 = 1/H^2$ from (8.4.25) (to leading order in slow roll), leads to

$$\mathcal{P}_T(k) = \frac{2H^2}{\pi^2\, M_{Pl}^2}\Big|_{\eta_0=-1/k_0}\Big(\frac{k}{k_0}\Big)^{n_T}\mathcal{T}_T(k)\,. \tag{8.4.30}$$

Therefore the tensor to scalar ratio, is given by

$$r(k) = \frac{\mathcal{P}_T(k)}{\mathcal{P}_\mathcal{R}(k)} = 16\epsilon_V(k_0)\Big(\frac{k}{k_0}\Big)^{4\epsilon_V - 2\eta_V}\,\frac{\mathcal{T}_T(k)}{\mathcal{T}_\mathcal{R}(k)}\,. \tag{8.4.31}$$

Thus we see that for non-Bunch-Davis initial conditions the single-field slow roll consistency condition are modified to

$$r(k_0) = -8n_T(k_0)\left[\frac{\mathcal{T}_T(k_0)}{\mathcal{T}_\mathcal{R}(k_0)}\right] \tag{8.4.32}$$

and the standard consistency condition is *not* fulfilled unless the transfer functions for curvature and tensor perturbations coincide at the pivot point k_0, which is obviously unlikely since the potentials for scalar and tensor perturbations are different and the pivot scale is arbitrary.

It remains to find $\mathcal{T}_\alpha(k)$. The mode equation (8.4.8) with $W(\eta)$ given by (8.4.12) can now be written as (we now drop the label α to avoid cluttering the notation)

$$\Big[\frac{d^2}{d\eta^2} + k^2 - \frac{\nu^2 - 1/4}{\eta^2}\Big]S(k;\eta) = V(\eta)S(k;\eta)\,, \tag{8.4.33}$$

and is converted into an integral equation via the retarded Green's function $G_k(\eta,\eta')$ obeying

$$\left[\frac{d^2}{d\eta^2} + k^2 - \frac{\nu^2 - \frac{1}{4}}{\eta^2}\right]G_k(\eta,\eta') = \delta(\eta-\eta') \quad ; \quad G_k(\eta,\eta') = 0 \text{ for } \eta' > \eta\,. \tag{8.4.34}$$

This Green's function is given by

$$G_k(\eta,\eta') = i\,[g_\nu(k;\eta)\;g^*_\nu(k;\eta') - g_\nu(k;\eta')\;g^*_\nu(k;\eta)]\,\Theta(\eta-\eta') \quad , \tag{8.4.35}$$

where $g_\nu(k;\eta)$ is given by eq.(8.4.19).

We are interested in obtaining the power spectra for wavelengths of cosmological relevance today which crossed the Hubble radius during slow roll inflation. These modes were deep inside the Hubble radius during the fast roll stage and we take these to be described by the asymptotic behavior of Bunch-Davies modes $\simeq e^{-ik\eta}/\sqrt{2k}$ for $-k\eta \gg 1$.

The solution of (8.4.33) with boundary conditions corresponding to Bunch-Davies modes deep inside the horizon during the fast roll stage obeys the following Lippman-Schwinger integral equation familiar from scattering theory,

$$S(k;\eta) = g_\nu(k;\eta) + \int_{\eta_i}^{0} G_k(\eta,\eta')\;V(\eta')\;S(k;\eta')\;d\eta'\,. \tag{8.4.36}$$

With the Green's function given by (8.4.35) this solution can be written as

$$S(k;\eta) = A_k(\eta) g_\nu(k;\eta) + B_k(\eta) g^*_\nu(k;\eta)\,, \tag{8.4.37}$$

where

$$A_k(\eta) = 1 + i\int_{\eta_i}^{\eta} V(\eta')\, g^*_\nu(k;\eta')\, S(k;\eta')\, d\eta' \tag{8.4.38}$$

$$B_k(\eta) = -i\int_{\eta_i}^{\eta} V(\eta')\, g_\nu(k;\eta')\, S(k;\eta')\, d\eta'\,. \tag{8.4.39}$$

In ref.[311] it is shown that the η dependent coefficients $A_k(\eta); B_k(\eta)$ obey

$$\frac{d}{d\eta}\left(|A_k(\eta)|^2 - |B_k(\eta)|^2\right) = 0 \tag{8.4.40}$$

which by dint of the initial conditions at η_i lead to the η-independent condition

$$|A_k(\eta)|^2 - |B_k(\eta)|^2 = 1\,. \tag{8.4.41}$$

Since the potentials vanish for $\eta > \eta_{sr}$, the solution of the mode equations during the slow roll stage is given by

$$S(k;\eta) = A_k g_\nu(k;\eta) + B_k g^*_\nu(k;\eta) \quad ; \quad \text{for} \quad \eta > \eta_{sr}\,, \tag{8.4.42}$$

namely of the form given by (8.4.18) with the Bogoliubov coefficients being the solutions of the integral equations

$$A_k = 1 + i\int_{\eta_i}^{\eta_{sr}} V(\eta')\, g^*_\nu(k;\eta')\, S(k;\eta')\, d\eta' \tag{8.4.43}$$

$$B_k = -i\int_{\eta_i}^{\eta_{sr}} V(\eta')\, g_\nu(k;\eta')\, S(k;\eta')\, d\eta' \tag{8.4.44}$$

where the potentials for curvature and tensor perturbations are given by (8.4.13,8.4.14) and (8.4.15,8.4.16) respectively. The quantity $|B_k|^2$ has the interpretation of the number of Bunch-Davies particles created by the potential during the fast roll stage.

Writing S_α as in eqn. (8.4.37) one obtains a coupled set of integral equations for the Bogoliubov coefficients and, following ref.[311], these can be written as a set of coupled differential equations which must be solved numerically in general. Ref.[311] provides an

analysis of the behavior of the Bogoliubov coefficients in the long-wavelength limit, valid for modes that are superhorizon well before the onset of the slow roll stage. These modes remain outside the current Hubble radius and are of no observational significance today. Instead, we focus on modes that are deep within the Hubble radius during the fast roll stage and cross during the slow roll stage.

The integral equations (8.4.43,8.4.44) can be solved formally as a Born series from the iterative solution of (8.4.36), namely

$$S(k;\eta) = g_\nu(k;\eta) + \int_{\eta_i}^{0} G_k(\eta,\eta')\, V(\eta')\, g_\nu(k;\eta')\, d\eta' + \cdots \,, \tag{8.4.45}$$

leading to the Born approximation for the Bogoliubov coefficients,

$$A_k = 1 + i\int_{\eta_i}^{\eta_{sr}} V(\eta')\, |g_\nu(k;\eta')|^2 d\eta' + \cdots \tag{8.4.46}$$

$$B_k = -i\int_{\eta_i}^{\eta_{sr}} V(\eta')\, \big(g_\nu(k;\eta')\big)^2 d\eta' + \cdots \tag{8.4.47}$$

where we have used that $V(\eta') = 0$ for $\eta > \eta_{sr}$.

Progress can be made by recognizing that we are interested in wavevectors that have crossed the horizon during slow roll since those are of cosmological relevance today, therefore these wavevectors are deep within the Hubble radius during the fast roll stage $\eta_i \le \eta \le \eta_{sr}$. The mode functions $g_\nu(k;\eta) \propto 1/\sqrt{k}$ for wavevectors deep inside the horizon, and from the expression for the potentials (8.4.13,8.4.15) and conformal time (8.4.17) we recognize that $V_{\mathcal{R},T} \propto H_{sr}^2$ and $\eta \propto 1/H_{sr}$ therefore we expect that $A_k - 1$; $B_k \propto H_{sr}/k$ suggesting that the lowest order Born approximation is reliable for wavectors that cross the horizon after η_{sr}, namely $k/H_{sr} > 1$ since we have normalized the scale factor so that $a(t_{sr}) = 1$. This expectation will be quantified and confirmed below. Furthermore to leading order in ϵ_V we will set $\nu = 3/2$ in the mode functions in (8.4.46,8.4.47) with (again up to a phase)

$$g_{3/2}(k;\eta) = -\frac{1}{\sqrt{2k}}\, e^{-ik\eta}\Big[1 - \frac{i}{k\eta}\Big] \,. \tag{8.4.48}$$

To leading order in ϵ_V and in the Born approximation, namely linear order in the potentials V we find

$$\mathcal{T}_\alpha(k) = 1 + \frac{1}{k}\int_{\eta_i}^{\eta_{sr}} V_\alpha(\eta)\left[\frac{2\cos(2k\eta)}{k\eta} + \sin(2k\eta)\Big(1 - \frac{1}{k^2\eta^2}\Big)\right] d\eta \,. \tag{8.4.49}$$

The potentials V_α have dimensions of H_{sr}^2 and η has dimensions of $1/H_{sr}$ therefore it is convenient to define the dimensionless functions of the variable x introduced in eqn. (8.2.30)

$$\tilde{\eta}(x) \equiv H_{sr}\eta(x)\,, \tag{8.4.50}$$

where $\eta(x)$ is given by (8.4.17) and

$$\widetilde{V}_\alpha(x) = \frac{V_\alpha(\eta(x))}{H_{sr}^2}\,, \tag{8.4.51}$$

along with the dimensionless ratio

$$q = \frac{k}{H_{sr}}\,, \tag{8.4.52}$$

in terms of which we find to leading order in the Born approximation

$$\mathcal{T}_\alpha(k) = 1 + D_\alpha(q) \tag{8.4.53}$$

with

$$D_\alpha(q) = \frac{1}{q}\left(\frac{12}{\epsilon_V}\right)^{1/6}\int_{x_i}^{x_{sr}} \frac{\widetilde{V}_\alpha(x)}{(1-x^6)^{1/3}}\left[\frac{2\cos(2q\tilde{\eta}(x))}{q\tilde{\eta}(x)} + \sin(2q\tilde{\eta}(x))\Big(1-\frac{1}{q^2\tilde{\eta}^2(x)}\Big)\right]dx\,. \tag{8.4.54}$$

The ratio q has a simple interpretation: assuming that during slow roll the Hubble parameter does not vary appreciably, namely $H \simeq H_{sr}(1+\mathcal{O}(\epsilon_V))$ at least during the range of the slow roll regime when wavevectors of relevance today crossed the Hubble radius, a comoving wavevector k corresponding to a physical scale that crosses the Hubble radius when the scale factor is $a_\star$ is given by $k = a_\star H_{sr}$ therefore $q = k/H_{sr} = a_\star$. Since we have normalized $a(t_{sr}) = a_{sr} = 1$ at the beginning of slow roll, values of $q > 1$ correspond to physical wavelengths that cross the Hubble radius during the slow roll stage. If the slow roll stage of inflation lasts about 60 e-folds the wavelengths of relevance today crossed out of the Hubble radius during the first few e-folds after the beginning of slow roll and modes with $q = a_\star > 1$ are of cosmological relevance today.

Figs.(39,40) show $D_R(q)$ and $D_T(q)$ for $\kappa = 10, 100$ for $\epsilon_V = 0.008, \eta_V = -0.01$. It is clear that the Born approximation is reliable for $q > 1$.

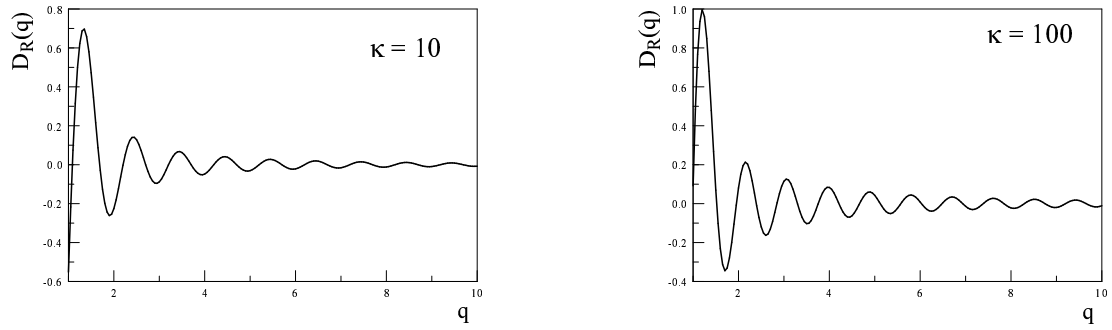


Figure 39: $D_R(q)$ vs. $q = k/H_{sr}$ for $\kappa = 10; 100$; $\epsilon_V = 0.008$; $\eta_V = -0.010$.

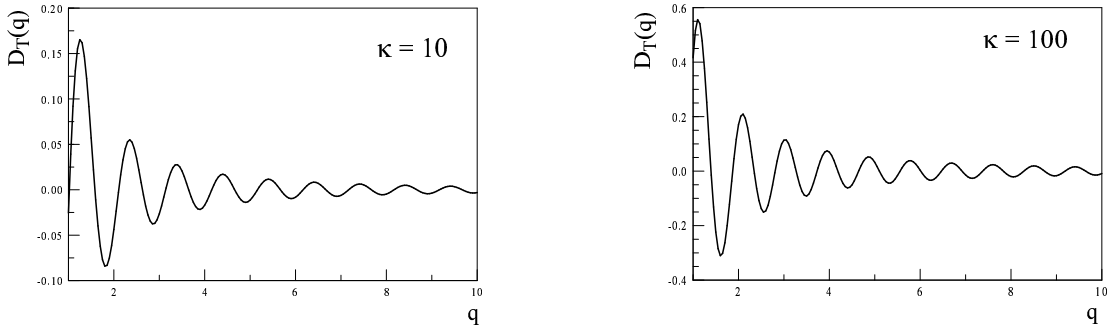


Figure 40: $D_T(q)$ vs. $q = k/H_{sr}$ for $\kappa = 10; 100$; $\epsilon_V = 0.008$; $\eta_V = -0.010$.

Although the Born approximation breaks down for $q < 1$ these figures confirm that the power spectra are suppressed at small q, as argued in ref.[275, 276], and feature oscillations of *the same frequency* determined by the Hubble scale during slow roll inflation as revealed by the figures (39,40). The observation of oscillations in the tensor to scalar ratio has been noted in other models, specifically in the double inflation model of ref.[316] in which two

distinct fields lead to separate periods of inflation leading to a mild oscillatory behavior in the period of transition.

Oscillatory behavior of the curvature and tensor power spectra and power suppression at small k was also observed in ref.[279] where a numerical integration of the mode equations from a kinetic dominated initial state was performed for the specific potential $\lambda\phi^4$. Although it is not straightforward to compare the scales, the discussion in this reference suggests that the oscillatory behavior is seen in modes that are very near horizon crossing and taper-off for larger values. This seems to be in agreement with our results that display oscillations for $q \simeq 1$ (namely $k \simeq H_{sr}$) but fall off as $\propto 1/q$ for $q >> 1$.

We emphasize that the results presented above depend solely on $\kappa, \epsilon_V; \eta_V$ but not on a specific realization of the inflationary potential, therefore are *universal* in this sense.

The relative change in the tensor to scalar ratio to leading order in the Born approximation is given by

$$\frac{\Delta r(k_0)}{r(k_0)} = D_T(q) - D_R(q) \tag{8.4.55}$$

where $q = k_0/H_{sr}$ and k_0 is the pivot scale. This relative change is displayed in fig.(41) for $\kappa = 10; 100$ for $\epsilon_V = 0.008; \eta_V = -0.010$.

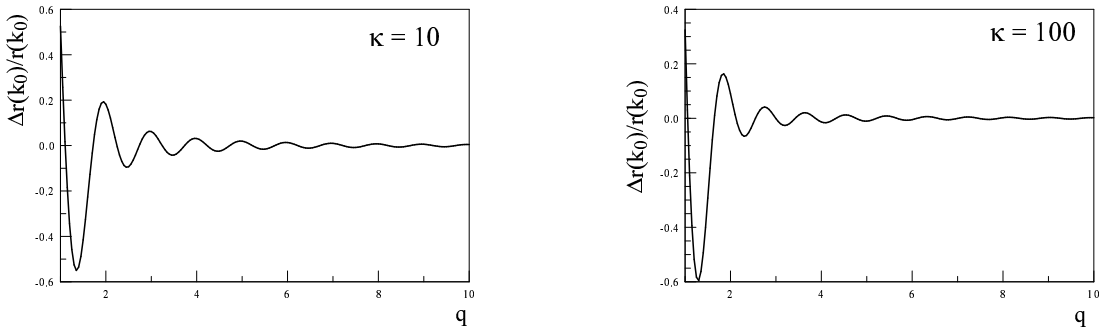


Figure 41: $\Delta r(k_0)/r(k_0) = D_T(q) - D_R(q)$ vs. $q = k_0/H_{sr}$ for $\kappa = 10; 100$; $\epsilon_V = 0.008$; $\eta_V = -0.010$.

The oscillatory features in the corrections to the curvature and tensor power spectra agree

qualitatively with the results obtained in refs.[275, 279] (which were obtained for different specific realizations of the inflationary potential). In fact, these oscillatory features are quite robust: the fast roll stage itself is insensitive to the potential, provided that the potential is very flat as is the case for slow roll inflation; it is only the merging with the slow roll stage that is sensitive to the potentials, but only through the slow roll parameters ϵ_V, η_V to leading order in the slow roll expansion. Thus different potentials that lead to the same slow roll parameters would yield the same type of behavior the functions D_α.

Correlation with suppression of low multipoles:

The modification on the initial conditions of the mode functions during slow roll imprinted from the pre-slow roll stage that lead to the corrections to the tensor to scalar ratio also affect the low multipoles in the CMB as previously discussed in refs.[273, 274, 275, 276, 306, 279]. In these references, specific inflationary potentials were studied whereas the analysis above, to leading order in the slow-roll parameters, is quite general and depends solely on $\kappa, \epsilon_V, \eta_V$. This allows us to study the modifications on the low multipoles in a more general manner in order to establish a correlation between features in the tensor to scalar ratio and the suppression of the low multipoles, in particular the quadrupole.

In the analysis that follows we neglect the contributions to the $C_l's$ from the integrated Sachs-Wolfe effect (the source of secondary anisotropies that could most likely affect the large scale anomalies in the CMB[250, 251]). This is supported by a recent analysis that suggests that a low quadrupole remains statistically significant and more anomalous even after subtraction of the (ISW) effect (see the discussion in ref.[251]).

The modifications from the non-Bunch-Davies initial conditions upon the temperature power spectrum are encoded in the transfer function $\mathcal{T}_R(k)$ and, to leading order in the Born approximation, by the correction $D_R(q)$ which is given by (8.4.54) for $\alpha = R$ and the potential is given by (8.4.14). This is depicted in fig.(37).

In the region of the Sachs-Wolfe plateau for $l \lesssim 30$ the matter-radiation transfer function can be set to one and, neglecting the contribution from the (ISW) effect through dark energy, the $C_l's$ are given by

$$C_l = \frac{4\pi}{9} \int_0^\infty \frac{dk}{k} \mathcal{P}_\mathcal{R}(k)\, j_l^2[k(\eta_0 - \eta_{lss})] \tag{8.4.56}$$

where the power spectrum $\mathcal{P}_{\mathcal{R}}(k)$ is given by (8.4.22) and

$$\eta_0 - \eta_{lss} = \frac{1}{a_0 H_0} \int_{1/(1+z_{lss})}^{1} \frac{dx}{\left[\Omega_r + \Omega_m x + \Omega_\Lambda x^4\right]^{1/2}} = \frac{3.12}{a_0 H_0} \tag{8.4.57}$$

where we have used $z_{lss} = 1100$ and the latest parameters reported by the Planck collaboration [63]. To leading order in the Born approximation we find the relative correction to the C_l from the initial conditions to be

$$\frac{\Delta C_l}{C_l} = \frac{\int_0^\infty dk k^{n_s - 2} D_R(k) j_l^2(3.12k/a_0 H_0)}{\int dk k^{n_s-2} j_l^2(3.12k/a_0 H_0)} . \tag{8.4.58}$$

In particular, taking $n_s = 1$ the corrections to the multipole l are[1]

$$\frac{\Delta C_l}{C_l} = 2l(l+1) \int_0^\infty \frac{dq}{q} D_R(q) j_l^2\left[\frac{3.12}{a_e} q\right] \tag{8.4.59}$$

where

$$a_e = \frac{a_0 H_0}{H_{sr}} \tag{8.4.60}$$

is the value of the scale factor when the physical scale corresponding to the Hubble radius today crossed the Hubble radius during slow roll inflation (with the normalization $a(t_{sr}) = a_{sr} = 1$ at the onset of slow roll).

The prominent oscillations in $D_R(q)$ do not yield oscillatory features in the ratio $\Delta C_2/C_2$ as a function of a_e. This result can be seen by combining (8.4.54) with (8.4.59), which leads to

$$\frac{\Delta C_2}{C_2} = \left(\frac{12}{\epsilon_V}\right)^{1/6} \int_{x_i}^{x_{sr}} \frac{\widetilde{V}_R(x)}{(1-x^6)^{1/3}} \Psi(\tilde{\eta}(x))\, dx \tag{8.4.61}$$

where the function $\Psi(\tilde{\eta}(x))$ has been studied in the second reference in[275] (see appendix of this reference), this function is non-oscillatory and positive for $\tilde{\eta} < 0$. Therefore for $V_R < 0$ it follows that ΔC_2 is non-oscillatory and negative as a function of a_e.

The relative change in the quadrupole and octupole are shown in fig. (42) for $\kappa = 10; 100$ and $\epsilon_V = 0.008; \eta_V = -0.010$.

These figures reveal that a $5 - 10\%$ suppression of the quadrupole, as reported by the Planck collaboration[300] can be accounted for if $a_e \simeq 2 - 3$. These results for $\Delta C_2/C_2$ are

[1]This expression corrects an overall normalization in the second reference in[275].

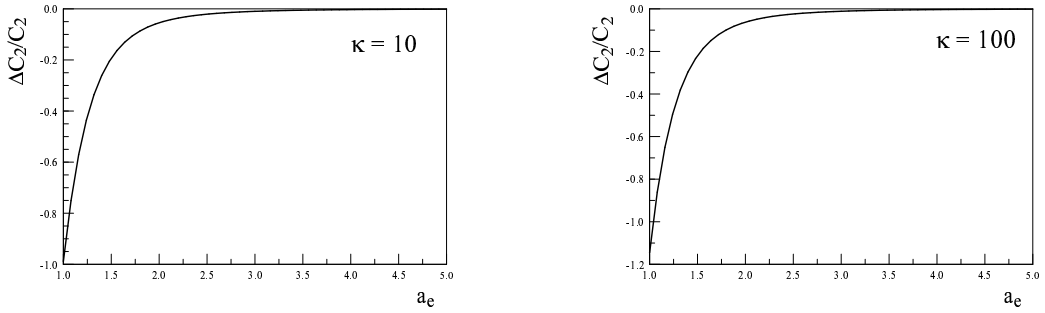


Figure 42: $\Delta C_2/C_2$ vs. a_e for $\kappa = 10; 100$; $\epsilon_V = 0.008$; $\eta_V = -0.010$.

qualitatively similar to those found in ref.[275] for a specific potential and different values of ϵ_V, η_V (as well as the argument of the Bessel function) and different initial conditions.[2]

The function $j_2^2\left[\frac{3.12}{z}\right]/z$ features a sharp peak at $z \simeq 1$ of width $\Delta z \simeq 1$, therefore the largest contribution to the integrand in (8.4.59) for the quadrupole ($l = 2$) arises from the region in q centered at $q \simeq a_e$ of width $\Delta q \simeq a_e$. From fig. (42) we see that for $2 \leq a_e \lesssim 4$ there is a suppression in the quadrupole in the range $0.05 \leq \Delta C_2/C_2 \lesssim 0.1 - 0.15$ which is approximately the suppression reported by the Planck collaboration[300]. Translating this range to fig. (41), we see that if the pivot scale k_0 is such that $2 \lesssim q = k_0/H_{sr} \lesssim 6 - 7$ then the tensor to scalar ratio should display oscillations with a period $\simeq H_{sr}$ as a function of the pivot scale.

Therefore, if the total number of inflationary e-folds is about the minimum for the scale corresponding to the Hubble radius today to have crossed the Hubble radius near the beginning of slow roll inflation, then the fast-roll stage would lead to a suppression of the quadrupole consistent with observations *and* oscillations in the tensor to scalar ratio. These *could* be observable if the wavelength corresponding to the pivot scale crosses the Hubble radius during slow roll just a few e-folds after the beginning of slow roll.

[2]There is also a normalization error in the second reference in[275].

Additionally, it has been pointed out in [314] that relieving the tension between Planck and BICEP by the invocation of a running spectral index is statistically less preferential than a mechanism which would lead to a large scale power suppression. While ref.[314] finds that a power suppression of $\simeq 35\%$ to be the best fit to Planck+BICEP data and we find a $\simeq 10\%$ suppression due to the fast roll stage, we emphasize that our analysis was only a leading order correction and further numerical effort would be needed to fully ascertain the effect of large scale power suppression due to a fast roll scenario.

We have also studied $\Delta C_l/C_l$ for $3 \leq l \leq 30$ in the Sachs-Wolfe plateau and found consistently that the higher multipoles are not substantially suppressed with respect to $\Delta C_2/C_2$. Fig. (43) displays the corrections to the octupole, which are consistently at least an order of magnitude smaller than $\Delta C_2/C_2$ in the whole range ($a_e > 1$), and displays a slight enhancement at $a_e \simeq 2$, a region where the quadrupole shows a suppression of $5-10\%$ which is consistent with Planck results[300]; however, the amplitude of such enhancement is $\simeq \mathcal{O}(\epsilon_V)$. For $l \geq 3$ the typical changes $\Delta C_l/C_l \ll \epsilon_V$ and, therefore, unobservable and indistinguishable from higher order corrections in ϵ_V.

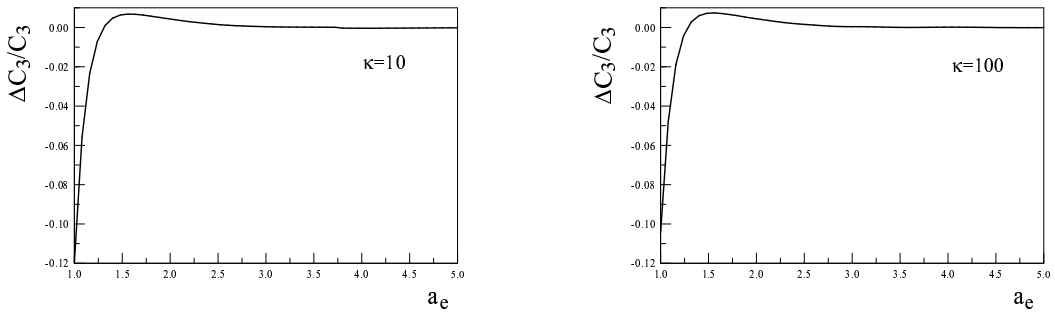


Figure 43: $\Delta C_3/C_3$ vs. a_e for $\kappa = 10; 100$; $\epsilon_V = 0.008$; $\eta_V = -0.010$.

Recently a method to search for oscillatory features in power spectra was introduced[315] that bears the promise of extracting important signatures from (CMB) data that could help to discern the effects of non-Bunch Davies initial conditions. In order to search for features

in (CMB) data it would be helpful to find a particular simple form for the corrections that could be useful for analysis advocated in ref.[315]. We find that while there is no simple fit to the damped oscillatory form of the corrections generally valid for all values of q, there are simple fits that are valid in a wide range of momenta of observational relevance. In particular, within the wide interval $1.5 \lesssim q \lesssim 10 - 15$, the following form is a very accurate fit both for curvature and tensor perturbations

$$D_\alpha(q) = \frac{A_\alpha(\kappa)}{q^{p(\kappa)}} \cos\left[2\pi\,\omega(\kappa)\,q + \varphi(\kappa)\right] \tag{8.4.62}$$

where the amplitude, power, frequency and phase-shift are slowly varying functions of κ within the wide range $3 \lesssim \kappa \lesssim 100$. Remarkably we find that the power and the frequency are *the same* for both types of perturbations, the power diminishes with κ within the range $1.5 \lesssim p(\kappa) \lesssim 2$ for $3 \lesssim \kappa \lesssim 100$ whereas $1 \lesssim \omega \lesssim 1.1$ within this range. These fits are shown in fig.(44) for both curvature and tensor perturbations. Whereas both the power p and frequency ω are the *same* (at least within the accuracy of the numerical fit) for both curvature and tensor perturbations, they differ both in amplitude and phase-shifts.

This analysis allows us to provide a compact form for the curvature and tensor power spectra that includes the modifications from the pre-slow roll stage and is valid within a wide range of observationally relevant momenta for the low-l region of the (CMB):

$$\mathcal{P}_R(k) = \mathcal{A}_R(k_0)\left(\frac{k}{k_0}\right)^{n_s-1}\left[1 + A_R(\kappa)\left(\frac{H_{sr}}{k}\right)^{p(\kappa)} \cos\left[2\pi\,\omega(\kappa)\,\frac{k}{H_{sr}} + \varphi_R(\kappa)\right]\right] \tag{8.4.63}$$

$$\mathcal{P}_T(k) = \mathcal{A}_T(k_0)\left(\frac{k}{k_0}\right)^{n_T}\left[1 + A_T(\kappa)\left(\frac{H_{sr}}{k}\right)^{p(\kappa)} \cos\left[2\pi\,\omega(\kappa)\,\frac{k}{H_{sr}} + \varphi_T(\kappa)\right]\right]. \tag{8.4.64}$$

Clearly there are further corrections to the Bunch-Davies part of the power spectra which are higher order in slow roll parameters ϵ_V, η_V, however, these are non-oscillatory and cannot mask the oscillatory contributions in the brackets of these expressions.

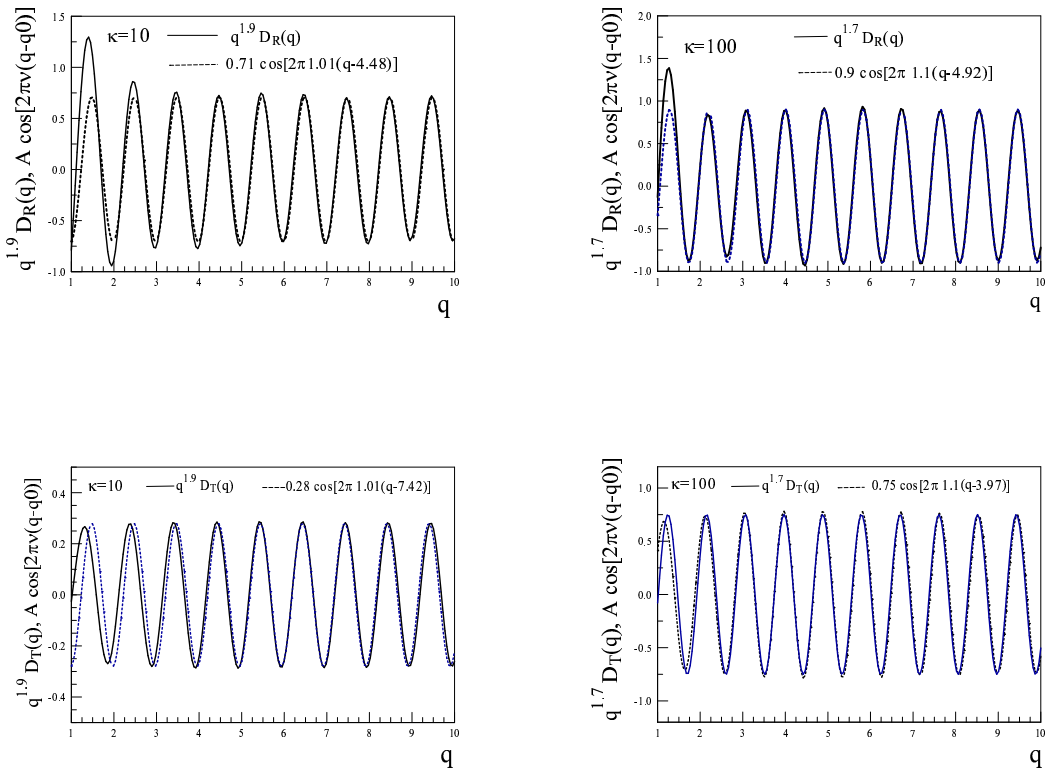


Figure 44: Fits to $D_{R,T}(q)$ within the range $1 \leq q \leq 10$ for $\kappa = 10; 100$; $\epsilon_V = 0.008$; $\eta_V = -0.010$. .

8.5 SUMMARY, CONCLUSIONS AND FURTHER QUESTIONS

Motivated by the most recent results from the PLANCK collaboration[63, 300, 301] reporting statistically significant anomalies at large scales, in this article we study the corrections to the curvature and tensor power spectra from non-Bunch-Davies initial conditions resulting from a fast-roll stage prior to slow roll in single field inflation with canonical kinetic terms. We consider initial conditions in which the kinetic energy of the inflaton is larger than the

potential energy, parametrized by the ratio

$$\frac{\dot{\Phi}_i^2}{2V_{sr}} = \kappa \gg 1 \tag{8.5.1}$$

where V_{sr} is the (fairly flat) inflaton potential consistent with slow roll. For a wide range of initial conditions with $\kappa \lesssim 100$ the pre-slow roll, kinetic dominated stage lasts for about $2-3$ e-folds with the inflationary stage beginning promptly within $\simeq 1$ e-folds, merging smoothly with the slow roll phase.

We have developed a program that yields the solution for the dynamics of the inflaton that interpolates between the fast and slow roll stages consistently in an expansion in ϵ_V, η_V which is *independent of the inflationary potential*. This approach relies on a separation of time scales and is valid for general inflationary potentials that are monotonic and can be described by a derivative expansion characterized by the slow roll parameters.

The fast roll stage modifies the potentials that enter in the equations of motion for the mode functions of curvature and tensor perturbations resulting in non-Bunch-Davies initial conditions for the mode functions *during slow roll*. The power spectra for curvature and tensor perturbations ($\alpha = R, T$) are modified to

$$\mathcal{P}_\alpha(k) = P_\alpha^{BD}(k)\mathcal{T}_\alpha(k) \tag{8.5.2}$$

where $\mathcal{T}_\alpha(k)$ are transfer functions determined by the non-Bunch-Davies initial conditions. These corrections entail a modification of the "consistency conditions" for the tensor to scalar ratio for single field inflation (with canonical kinetic term) to

$$r(k_0) = -8n_T(k_0)\left[\mathcal{T}_T(k_0)/\mathcal{T}_\mathcal{R}(k_0)\right] \tag{8.5.3}$$

where k_0 is the "pivot" scale.

We obtain explicit expressions for the $\mathcal{T}_\alpha(k)$ in a Born approximation which is valid for all modes of observational relevance today; i.e. those that had crossed the Hubble radius during inflation within $1-2$ e-folds from the beginning of the slow roll stage. The modification of the power spectrum for curvature perturbations yields a suppression of the (CMB) quadrupole consistent with the results from Planck[300] if the modes corresponding to the Hubble radius

today crossed the Hubble radius within a few ($\simeq 2-3$) e-folds from the beginning of slow roll, suggesting that a kinetic dominated pre-slow roll stage is a possible explanation of the quadrupole suppression if the number of e-folds during slow roll inflation is the minimal required to solve the horizon problem $\simeq 60-62$. As discussed in [314], a large scale power suppression is a mechanism which could serve to relieve the tension between the Planck and BICEP experiments which would imply that a fast roll stage *could* potentially serve as a mechanism to explain the seemingly conflicted results.

The suppression of the quadrupole is correlated with oscillatory features in the tensor to scalar ratio which *could* be observable, again if the modes corresponding to the observed pivot scale crossed the Hubble radius a few e-folds from the beginning of slow roll inflation.

A numerical fit to the power spectra valid for these wavevectors yields

$$\mathcal{P}_R(k) = \mathcal{A}_R^{BD}(k_0)\left(\frac{k}{k_0}\right)^{n_s-1}\left[1+A_R(\kappa)\Big(\frac{H_{sr}}{k}\Big)^{p(\kappa)}\cos\left[2\pi\,\omega(\kappa)\,\frac{k}{H_{sr}}+\varphi_R(\kappa)\right]\right] \tag{8.5.4}$$

$$\mathcal{P}_T(k) = \mathcal{A}_T^{BD}(k_0)\left(\frac{k}{k_0}\right)^{n_T}\left[1+A_T(\kappa)\Big(\frac{H_{sr}}{k}\Big)^{p(\kappa)}\cos\left[2\pi\,\omega(\kappa)\,\frac{k}{H_{sr}}+\varphi_T(\kappa)\right]\right] \tag{8.5.5}$$

remarkably with the same power $p(\kappa)$ and frequency $\omega(\kappa)$ for *both* tensor and curvature perturbations, with H_{sr} the Hubble scale during inflation and

$$1.5 \lesssim p(\kappa) \lesssim 2 \; ; \; \omega(\kappa) \simeq 1 \tag{8.5.6}$$

and

$$0.7 \lesssim A_R(\kappa) \lesssim 0.9 \;\; ; \;\; 0.3 \lesssim A_T(\kappa) \lesssim 0.8 \tag{8.5.7}$$

for the range $3 \lesssim \kappa \lesssim 100$.

Perhaps these oscillatory features in the power spectra *may be* extracted from (CMB) data with the implementation of the techniques advocated recently in ref.[315]. The Bunch-Davis contribution to the power spectra will feature higher order corrections in ϵ_V, η_V which *may be comparable in magnitude* to the corrections brought about by the fast roll phase, however, the distinguishing oscillatory features in the power spectra above cannot be confused with the non-oscillatory higher order slow roll corrections.

The results discussed above were obtained within the regime of validity of the Born approximation and to leading order in an expansion in ϵ_V, η_V, therefore there remains the question of possible corrections beyond this approximation, for which a more definitive answer would imply either extending the calculation to higher orders in the Born series and the ϵ_V, η_V expansion of section (8.3) or a full numerical solution of the mode equations and the Friedmann equation. Both approaches imply a substantial and intensive numerical effort, an endeavor that would be justified if the analysis of the (CMB) data yields hints of oscillatory behavior in broad agreement with the scales and general features of the results of the leading order approximation described by the power spectra above. With the recent detection of primordial B-waves in ref. [302], the proposal of scanning across pivot scales in an effort to observe the aforementioned oscillations in the tensor to scalar ratio is a potentially realistic future goal and, if such oscillations are detected, direct access to pre-inflationary information may be within the realm of plausibility.

9.0 COSMOLOGICAL IMPLICATIONS OF LIGHT STERILE NEUTRINOS PRODUCED AFTER THE QCD PHASE TRANSITION

Based on: (ref. [396])

L. Lello, D. Boyanovsky, Phys. Rev. D 91, 063502 (2015)

9.1 INTRODUCTION

The current paradigm in cosmology is that the energy content of the universe is divided into the particle species of the standard model, an unknown dark energy driving the current expansion of the universe and an unknown (cold) dark matter species (ΛCDM) [317]. Dark matter (DM) is thought to be in the form of cold thermal relics with interaction cross sections on the order of weak interaction strength (WIMPs) [318] with alternate theories favoring axions [319] or new neutrino species [320]. The standard cold dark matter cosmology explains much of the observational data yet some problems at small scales remain unexplained.

Cold dark matter N body simulations predict that dark matter dominated galaxy profiles feature a cusp, but observations suggest that the profiles are cores (core v cusp problem) [321, 322]. Additionally, simulations of ΛCDM show that dark matter subhaloes in the Milky way are too dense for the observed satellites (too big to fail) [323]. Both of these problems could be alleviated if the dark matter candidate is allowed to be "warm" (WDM) [324, 325, 326, 327, 328, 329], one such candidate being a massive "sterile" neutrino [330, 69, 331, 70, 332]. The free streaming length, $\lambda_{fs} = 2\pi/k_{fs}$, is the scale that cuts off the

power spectrum of density perturbations. CDM features very small ($\lesssim$ pc) λ_{fs} which leads to cuspy profiles while WDM features $\lambda_{fs} \sim$ few kpc possibly explaining the observed cores. λ_{fs} is determined by the distribution function at freeze out. Alternatively, decaying DM candidates, such as WIMPs or gravitinos, could also lead be a simultaneous solution to both of these problems [333].

Additionally, with the discovery of neutrino masses, a considerable experimental effort has shed light on the parameters of the neutrino sector [334, 335]. The last of the mixing angles describing neutrino oscillation has been measured and there are proposals for new facilities to probe CP violation, Dirac/Majorana nature, inverted/normal hierarchy in the active neutrino sector [336]. There are also some persistent short baseline anomalies (LSND, MiniBooNE) [43, 45] that can be explained with an additional sterile neutrino species [332] but tension exists with other experiments [337]. There are plans to search for these sterile neutrinos in forthcoming experiments, many of which involve neutrino production from the decay of meson parent particles, processes in which the subtleties of the decay event itself may prove useful [338]. Other proposed experiments could search for sterile neutrinos via modifications to oscillation formulae on short baseline experiments [139] , monochromatic peaks searches [139, 89] or as contributions to lepton flavor violation experiments [339]. A review of the motivation for sterile neutrinos from terrestrial experiments and a summary of some of the proposed experiments that will look for sterile neutrinos can be found in [340]. The latest limits on sterile neutrino mixing from atmospheric neutrino data have been set by the Super Kamiokande experiment [341] which sets the limits $|U_{\mu 4}|^2 < 0.041$. Similar bounds have been by the Daya Bay collaboration [337] and the analysis in [343, 342] examines the global fits for various light sterile neutrino scenarios (3+1,3+2,3+1+1). A summary of the light sterile neutrino bounds for active-sterile mixing from accelerators, cosmology and other experiments are summarized concisely in figs 1-3 of ref [71] while those for heavy steriles can be found in [98].

Several extensions of the standard model include sterile neutrino species, for instance [344] describes a model which is an extension of the νMSM and purports to describe inflation, dark matter, the baryon asymmetry and neutrino oscillations. For most treatments of sterile neutrino dark matter, a nonthermal distribution function is needed in order to

evade cosmological bounds [63]. Ref [345] argues that short baseline inspired steriles (1eV) could not be in thermal equilibrium in the early universe but can be made compatible with observations by allowing the sterile to decay into very light particles. The mechanism of sterile neutrino production in the early universe through oscillations was originally studied in a body of work by Barbieri, Dolgov, Enqvist, Kainulainen and Maalampi (BDEKM) [346] and, in [69], sterile neutrinos are argued to be a viable warm dark matter candidate produced out of LTE via the BDEKM mechanism (Dodelson-Widrow, DW). In [347], light keV sterile neutrinos are produced by resonant MSW conversion of active neutrinos, similiar to DW but with resonant oscillation in the presence of a lepton number asymmetry (Shi-Fuller, SF). Models in which a standard model Higgs scalar decays into pairs of sterile neutrinos at electroweak energy scales (or higher) have also received attention [348, 349, 350]. Ref [350] calculates the free streaming length and phase space density of sterile neutrinos from Higgs-like decays, both in and out of equilibrium, which is used to compare to small scale structure observations. These types of mechanisms have inspired work on understanding properties of more general nonthermal dark matter such as [74, 426].

Recently, a signal of 3.5 keV line has been claimed at 3σ detection from the XMM Newton x-ray telescope which could be a hint of a 7 keV sterile neutrino [351, 352]. The interpretation of the anomalous line as a signal of a sterile neutrino has been challenged [353, 354] motivating further studies of the signal. In refs [355, 70], the parameter space for SF type steriles that could be compatible with the 3.5 kev signal is explored. Besides the 3.5 keV line, other observational clues seem to favor or disfavor the various mechanisms. Ref [356] claims that high redshift quasar Ly α signals disfavor both DW and SF mechanisms but is consistent with scalar decay. Radiative decays of sterile neutrino dark matter candidates is constrained by the Chandra X-ray spectrum which places limits on sterile mass (for DW) at $m < 2.2keV$ [393]. Observations of dwarf spheroidal phase space densities and X-ray data in the local group essentially rule out DW steriles but still allow for SF or other mechanisms [357]. The effects of massive neutrinos on the Sachs Wolf plateau and CMB fluctuations have been calculated and limits placed on the mass and lifetime [358] while phase space densities of dwarf spheroidals lead to bounds a WDM sterile candidate at $m \lesssim$ few keV [359].

The prospect of keV WDM sterile neutrinos remains an active area of investigation

experimentally and theoretically. Ref [349] claims keV neutrino DM produced via Higgs decays matches the bounds of small scale structure and X-ray observations while simultaneously explaining pulsar kicks. It has also been suggested that SF type steriles reproduce the appropriate galaxy distribution and could potentially lead to a test of the quark-hadron transition [355]. Ranges of masses and mixing for both DW and SF mechanisms include constraints from supernovae, BBN and decay limits which can be found in [70]. One of the observational windows towards the detection of light ($m \lesssim$ eV) sterile neutrinos are from cosmological measurements of N_{eff}, the sum of neutrino masses and the lepton asymmetry and BBN [360] [361]. A comparison of how various dark radiation sources contribute to these measurements can be found in [362].

Ref [363] considers heavy sterile neutrinos (100-500 MeV) in thermal equilibrium but decay nonthermally and finds a range of parameter space in which these models can contribute to N_{eff} without violating the bounds. A mechanism of neutrino reheating in ref [364] considers other particles which remain in local thermodynamic equilibrium (LTE) with neutrinos and decouple before photon decoupling, changing the neutrino to photon temperature ratio. Contributions to N_{eff} from decaying non-thermal particles can mimic sterile neutrinos where higher moments of the distribution functions would be required to discriminate between scenarios [365]. Its been shown that delaying neutrino freeze out contributes to dark radiation [366] and, additionally, freeze out of Bose or Fermi degrees of freedom during QCD phase transition would lead to changes in dark radiation measurements [367]. Additionally, the relic densities of sterile neutrinos depend on the QCD transition and, if detection and study of these particles were possible, could offer a window to the QCD phase transition [70].

To the best of our knowledge, the mechanism which is used to produce neutrinos in many terrestrial experiments, $\pi \rightarrow \mu\nu$, has not been addressed in a cosmological setting. The difficulty in such a problem is reflected in the challenges inherent to the QCD era of the early universe. The QCD phase transition, when the universe cools enough for free quarks and gluons to hadronize, continues to be an epoch in cosmology which remains to be fully understood [368]. Recently, the latest lattice QCD calculations have suggested that the QCD phase transition is continuous with a crossover at $T = 155MeV$ [369]. It is generally accepted that π mesons, the lowest lying QCD bound states, will be produced in abundance

and this has motivated thorough studies of pions near the QCD phase transition. Near the phase transition, stable long wavelength pion excitations are developed which may be detectable signatures in heavy ion colliders [370, 371]. At temperatures below the QCD phase transition, finite temperature corrections to the pion mass and decay constant become important and non-trivial [372]. These corrections have been studied in linear [373] and non-linear sigma models[374], using QCD sum rules at finite temperature [375], hidden local symmetry models [376] and chiral perturbation theory [377].

Goals: The main goal of this work is to understand the production and freeze out of sterile neutrinos from $\pi \to l\nu_s$ shortly after the QCD transition. With the finite temperature corrections to the pion mass and decay constant, it is possible to consider the quantum kinetics of sterile neutrinos that are produced in the early universe from the same mechanisms which are employed by land based accelerator experiments, namely $\pi \to l\nu_s$. We obtain the distribution function of a sterile neutrino produced from pion decay in the early universe by including finite temperature corrections and investigate the immediate observational consequences. We will be restricting our attention to the study of *light* sterile neutrinos with masses $m_\nu \lesssim 1MeV$. These will be shown to freeze out while they are still relativistic with non-thermal distributions.

- With a non-thermal distribution function, measurements of Ω_{DM} give an upper bound for the energy density of the sterile neutrinos today. A complementary bound is obtained by considering the velocity dispersion and energy density of dwarf spheroidal galaxies. These measurements coupled with the non thermal distribution place bounds on combinations of masses and mixing matrix.
- The free streaming length, which is small for cold dark matter candidates and larger for warmer dark matter candidates, is dependent on the specific form of the distribution function. We obtain λ_{fs} from the non-thermal distribution function arising from pion decay.
- A light sterile neutrino of $m \lesssim 1$ eV could be relativistic at the time of matter-radiation equality and potentially contribute to the measurement of N_{eff}. We investigate the

contribution to this number from the pion-produced sterile neutrino and how the equation of state parameter, w, evolves from relativistic to non-relativistic compared to a thermal distribution.

Brief Summary of Results:

- We find the *non-thermal* distribution function for sterile neutrinos that were produced via pion decays shortly after the QCD phase transition. This distribution features a low momentum enhancement similar to that found in resonantly produced models (Shi-Fuller). A key difference between the two models is that resonant model requires a non-zero lepton asymmetry which is absent in the distribution that we obtain. This mechanism produces a colder sterile neutrino dark matter candidate, similar to MSW resonance enhancement, but without the requirement of a lepton asymmetry. A calculation of the equation of state shows that, while freeze-out occurs as the particles are still relativistic, this type of sterile neutrino becomes non-relativistic very quickly, namely when $T \sim m$, as opposed to thermal distributions which become non-relativistic when $T \ll m$.
- We obtain bounds on combinations of sterile neutrino mass and mixing matrix elements from CMB observations and dark matter dominated galaxies. Using the observed dark matter density from Planck as an upper bound for the sterile neutrino energy density leads to an upper bound on a combination of the mass and mixing matrix:

$$m_{\nu_s}\frac{|U_{\mu s}|^2}{10^{-5}} \leq 0.739\,\text{keV} \quad ; \quad m_{\nu_s}\frac{|U_{es}|^2}{10^{-5}} \leq 7242\,\text{keV}\,. \tag{9.1.1}$$

A complementary bound is obtained from the primordial phase space density and compared to present day observations of dark matter dominated galaxies. By requiring that the primordial phase space density of sterile neutrinos be larger than the observed density and velocity dispersion relations for dark matter dominated galaxies leads to a lower bound on a different combination of mass and mixing matrix:

$$m_\nu\left(\frac{|U_{\mu s}|^2}{10^{-5}}\right)^{1/4} \geq 0.38\,\text{keV} \quad ; \quad m_\nu\left(\frac{|U_{es}|^2}{10^{-5}}\right)^{1/4} \geq 6.77\,\text{keV}\,. \tag{9.1.2}$$

The 7.1 keV sterile neutrino predicted by [351, 352] (with $|U|^2 = 7 * 10^{-11}$) is consistent with these bounds for sterile neutrinos produced from $\pi \to \mu\nu_s$ within a narrow region but not from $\pi \to e\nu_s$.

- To be a suitable dark matter candidate, the free streaming length must be smaller than the size of the dark matter halo. The free streaming length is calculated using the non thermal distribution function and, due to the enhancement at low momentum, is reduced for keV type steriles. The free streaming length today is given by

$$\lambda_{fs}^{\mu}(0) \sim 7.6\,\mathrm{kpc}\left(\frac{\mathrm{keV}}{m_\nu}\right) \quad ; \quad \lambda_{fs}^{e}(0) \sim 16.7\,\mathrm{kpc}\left(\frac{\mathrm{keV}}{m_\nu}\right) \tag{9.1.3}$$

A sterile species that is still relativistic at the time of matter-radiation equality will contribute to N_{eff} and, since this type of sterile neutrino becomes non-relativistic at $T \sim m$, the contributions to N_{eff} are only valid for $m_\nu \lesssim 1eV$. Parameterizing the contribution to dark radiation as $N_{eff} = N_{eff}^0 + \Delta N_{eff}$ where $N_{eff}^0 = 3.046$ is the standard model contribution [378], the sterile neutrinos we consider here contribute

$$\Delta N_{eff}\Big|_{\pi\to\mu\nu} = 0.0040 * \frac{|U_{\mu s}|^2}{10^{-5}} \quad ; \quad \Delta N_{eff}\Big|_{\pi\to e\nu} = 9.7 * 10^{-7}\frac{|U_{es}|^2}{10^{-5}}\,. \tag{9.1.4}$$

Combining with a recent analysis [337, 341] we find that $\Delta N_{eff} \lesssim 4$, suggesting that this mechanism could provide a significant contribution to N_{eff} although severe tensions remain between accelerator/reactor fits and CMB observations.

9.2 DYNAMICS OF DECOUPLED PARTICLES

In this section we gather the general essential ingredients for several cosmological quantities in terms of the distribution function of the dark matter particle. Kinetic theory in a cosmological setting is well understood [379, 380, 381], the purpose of this section is to review the details of the dynamics of decoupled particles which will be relevant for the following sections. The results of this section will be used in conjunction with the distribution obtained from quantum kinetics to place limits on sterile neutrino parameters.

For flat Friedmann-Robertson-Walker (FRW) cosmologies, particles follow geodesics described by

$$ds^2 = dt^2 - a(t)^2 d\vec{x}^2 \,. \tag{9.2.1}$$

The only non-vanishing Christoffel symbols are given by

$$\Gamma^i_{j0} = \Gamma^i_{0j} = \frac{\dot{a}}{a}\delta^i_j \quad ; \quad \Gamma^0_{ij} = \dot{a}a\delta_{ij} \,. \tag{9.2.2}$$

The geodesic equations are then given by

$$\dot{q}^0 = -\frac{a^2 H \vec{q}^2}{q^0} \quad ; \quad \dot{\vec{q}} = -2H\vec{q} \tag{9.2.3}$$

where $q^\mu = dx^\mu/d\lambda$ and λ is an affine parameter. The solution is given by

$$\vec{q} = \frac{\vec{q}_c}{a^2} \tag{9.2.4}$$

where $\vec{q}_c$ is a constant comoving momentum. The geodesics of massive particles imply $g_{\mu\nu}q^\mu q^\nu = m^2$, leading to the dispersion relation $q^0 = \sqrt{m^2 + a^2\vec{q}^2}$.

The physical energy and momentum is that which is measured by an observer at rest with respect to the expanding spacetime. The stationary observer is one who measures with an orthormal tetrad

$$g_{\mu\nu}\varepsilon^\mu_\alpha \varepsilon^\nu_\beta = \eta_{\alpha\beta} = diag(1,-1,-1,-1) \tag{9.2.5}$$

or

$$\varepsilon^\mu_\alpha = \sqrt{|g^{\mu\alpha}|} \,. \tag{9.2.6}$$

With this, the physical energy/momentum are given by

$$E = g_{\mu\nu}\varepsilon^\mu_0 q^\mu = q^0 \quad ; \quad Q_f = g_{\mu\nu}\varepsilon^\mu_i q^\nu = aq^i = \frac{q^i_c}{a} \,. \tag{9.2.7}$$

The buildup of the distribution function arises from a Boltzmann equation in which decaying particles source the equation. Provided that any other interactions can be neglected, such as a sterile neutrino's interaction with standard model particles, and that the distribution is isotropic, then the kinetic equation is given by

$$\frac{df}{dt}(Q_f, t) = \frac{\partial f}{\partial t} - HQ_f \frac{\partial f}{\partial Q_f} = \mathbb{P}[f] \tag{9.2.8}$$

where $\mathbb{P}$ is the production integral which will be discussed in a subsequent section. Upon freeze out, the production integral vanishes and the distribution function follows geodesics governed by a collisionless Liouville equation, namely with $\mathbb{P} = 0$. We denote the decoupled distribution as f_d to distinguish it from the full distribution which is explicitly a function of time. It is easy to see that a solution for the decoupled distribution (with $\mathbb{P} = 0$) are functions of the form

$$f_d(Q_f, t) = f_d(a(t)Q_f) = f_d(q_c) \tag{9.2.9}$$

which depends on the scale factor through the comoving momentum.

For this type of distribution function, not necessarily thermal, the kinetic stress-energy tensor is given by

$$T^\mu_\nu = g \int \frac{d^3Q_f}{(2\pi)^3} \frac{q^\mu q_\nu}{q^0} f_d(q_c) \tag{9.2.10}$$

where g is the internal degrees of freedom of the particular species. The number density, energy density and pressure are obtained in a straightforward manner as

$$n = g \int \frac{d^3Q_f}{(2\pi)^3} f_d(q_c) \quad ; \quad \rho = T^0_0 = g \int \frac{d^3Q_f}{(2\pi)^3} \sqrt{Q_f^2 + m^2} f_d(q_c) \tag{9.2.11}$$

$$T^i_j = -\delta^i_j \frac{g}{3} \int \frac{d^3Q_f}{(2\pi)^3} \frac{|\vec{q}_c|^2}{E_q} f_d(q_c) \quad \rightarrow \quad \mathcal{P} = \frac{g}{3} \int \frac{d^3Q_f}{(2\pi)^3} \frac{|\vec{Q}_f|^2}{\sqrt{Q_f^2 + m^2}} f_d(q_c)\,. \tag{9.2.12}$$

Then, introducing the photon energy density today, we can write the contribution to the energy density as

$$\Omega h^2 = \frac{\rho h^2}{\rho_{crit}} = \frac{h^2 n_\gamma}{\rho_c} \frac{\pi^2 \rho}{2\zeta(3) T_\gamma^3} \,. \tag{9.2.13}$$

The average momentum squared per particle is given by

$$\overline{\vec{Q}^2} = \frac{\int \frac{d^3 Q_f}{(2\pi)^3} \vec{Q}_f^2 f_d(q_c)}{\int \frac{d^3 Q_f}{(2\pi)^3} f_d(q_c)} \,. \tag{9.2.14}$$

For a nonrelativistic species this is related to the average velocity per particle via $\overline{\vec{Q}^2} = m^2 \overline{\vec{V}^2}$ and to the pressure/energy density as will be discussed shortly. The Hubble factor in a radiation-dominated cosmology is given by

$$H(t) = 1.66 \, \frac{g(T)^{1/2} T(t)^2}{M_p} \,. \tag{9.2.15}$$

Since the distribution function after freeze-out obeys the Liouville equation, it is straightforward to verify that the number density and energy density obey a continuity equation

$$\frac{dn}{dt} + 3H(t) n(t) = 0 \quad ; \quad \frac{d\rho}{dt} + 3H(t)(\rho(t) + \mathcal{P}(t)) = 0 \,. \tag{9.2.16}$$

The entropy density for an arbitrary distribution function is given by

$$s_d(t) = -g \int \frac{d^3 q_f}{(2\pi)^3} \Big[f_d \ln f_d \pm (1 \mp f_d) \ln(1 \mp f_d) \Big] \tag{9.2.17}$$

where the upper (lower) is for fermions (bosons). For frozen distribution functions, ie one obeying a collisionless Liouville equation, we have a another continuity equation

$$\frac{ds}{dt} + 3H(t) s(t) = 0 \,. \tag{9.2.18}$$

This gives the result that the comoving entropy density, sa^3, is constant.

With a mixture of several types of species in LTE and additional non-thermal species with entropy s_d, entropy conservation gives

$$\left[\frac{2\pi^2}{45}g(T)T_\gamma^3 + s_d\right]a^3(t) = const\,. \tag{9.2.19}$$

where T_γ is the photon temperature and

$$g(T) = \sum_{i=Bosons} g_i\left(\frac{T_i}{T_\gamma}\right)^3 + \frac{7}{8}\sum_{j=Fermions} g_j\left(\frac{T_j}{T_\gamma}\right)^3 \tag{9.2.20}$$

where $T_{i/j}$ are the temperatures of the individual relativistic species. Since the non thermal particles obey $sa^3 = const$ the standard $g(T)a(T)^3T_\gamma^3 = const$ still holds even in the presence of non-thermal species (assuming instantaneous reheating of the photon gas when species give off entropy upon annihilation), namely

$$\frac{T_d(t)}{T_\gamma(t)} = \left(\frac{2}{g_d}\right)^{1/3} \rightarrow T_d(t) = \left(\frac{2}{g_d}\right)^{1/3} T_{\gamma,0}(1+z) \tag{9.2.21}$$

where T_d, g_d are the temperature and effective degrees of freedom at decoupling and $T_{\gamma,0}$ is the CMB temperature today.

Choosing the normalization $a_{today} = 1$, the temperature evolves as $T(t) = T_0/a(t)$, where T_0 is the temperature of the plasma today ($T_0 = (2/g_d)^{1/3}T_{\gamma,0}$) we can rewrite the density and pressure by introducing the dimensionless quantities, $x = m/T(t), y = q_f(t)/T(t) = q_c/T_0$ to give

$$\rho = \frac{gm}{2\pi^2}T^3(t)\left\langle y^2\sqrt{1+\frac{y^2}{x^2}}\right\rangle \quad ; \quad \mathcal{P} = \frac{g}{6\pi^2 m}T^5(t)\left\langle \frac{y^4}{\sqrt{1+\frac{y^2}{x^2}}}\right\rangle$$
$$\langle g(x,y)\rangle \equiv \int dy\, g(x,y)f_d(y) \tag{9.2.22}$$

where we've introduced the definition of $\langle g(x,y)\rangle$. Then the equation of state parameter is

given by

$$w = \frac{\mathcal{P}}{\rho} = \frac{1}{3x^2} \frac{\left\langle \frac{y^4}{\sqrt{1+\frac{y^2}{x^2}}} \right\rangle}{\left\langle y^2 \sqrt{1+\frac{y^2}{x^2}} \right\rangle} \tag{9.2.23}$$

For non-relativistic species $x \gg 1$ so we neglect the $(y/x)^2$ terms to arrive at the familiar result

$$\begin{aligned} \rho_{nr} = m\frac{g}{2\pi^2}T^3(t)\langle y^2\rangle = mn(t) \quad &; \quad \mathcal{P}_{nr} = \frac{g}{6\pi^2 m}T^5(t)\langle y^4\rangle \\ w_{nr} = \frac{T(t)^2}{3m^2}\frac{\langle y^4\rangle}{\langle y^2\rangle} \quad &\rightarrow \quad 0\,. \end{aligned} \tag{9.2.24}$$

For relativistic species, $x \ll 1$ and $\mathcal{P}_{rel} = \rho_{rel}/3$. Explicitly, the thermodynamic quantities become

$$\begin{aligned} \rho_{rel} = \frac{g}{2\pi^2}T^4(t)\langle y^3\rangle \quad &; \quad \mathcal{P}_{rel} = \frac{g}{6\pi^2}T^4(t)\langle y^3\rangle \\ w_{rel} = \frac{\langle y^3\rangle}{3\langle y^3\rangle} \quad &= \quad \frac{1}{3}\,. \end{aligned} \tag{9.2.25}$$

In the non relativistic limit, the average velocity per particle is given by

$$\overline{\vec{V}^2} = \frac{\overline{\vec{Q}^2}}{m^2} = \frac{T(t)^2}{m^2}\frac{\langle y^4\rangle}{\langle y^2\rangle} = \frac{3\mathcal{P}}{\rho} \tag{9.2.26}$$

which leads to the velocity dispersion relation

$$\mathcal{P} = \sigma^2\rho \;\; ; \;\; \sigma = \sqrt{\frac{\overline{\vec{V}^2}}{3}} = \frac{T(t)}{m}\sqrt{\frac{\langle y^4\rangle}{3\langle y^2\rangle}} \tag{9.2.27}$$

The work of Tremaine and Gunn [382] and Lynden-Bell [383] argued that the phase space density may only decrease as a galaxy evolves (violent relaxation and phase mixing). The phase space density is related to observationally accessible quantities (galactic velocity dispersion and density), and therefore the primordial phase space density can be used as an

upper bound to place limits on dark matter parameters. For dwarf galaxies, these observations are summarized in [359] and the phase space density is given by

$$\mathcal{D} = \frac{n(t)}{\vec{Q}^{2\,3/2}}\,. \tag{9.2.28}$$

The phase space density is completely determined by *moments* of the distribution function after freezeout. In terms of an arbitrary distribution function, this is given by

$$\mathcal{D} = \frac{g}{2\pi^2}\frac{\langle y^2\rangle^{5/2}}{\langle y^4\rangle^{3/2}} \tag{9.2.29}$$

During galactic evolution, the phase space decreases from its primordial value [322, 383]. Eventually, today, the particles will be non relativistic and, for a non relativistic particle, we have that

$$\mathcal{D}_{nr} = \frac{n}{\overline{Q_f^2}^{3/2}} = \frac{\rho}{m^4\overline{\vec{V}^2}^{3/2}} = \frac{1}{3^{3/2}m^4}\frac{\rho}{\sigma^3} \tag{9.2.30}$$

For a primordial phase space density, $\mathcal{D}_p$, imposing the bound $\mathcal{D}_p \geq \mathcal{D}_{nr}$ gives us the constraint

$$\mathcal{D}_p \geq \frac{1}{3^{3/2}m_{\nu_s}^4}\left.\frac{\rho}{\sigma^3}\right|_{today} \tag{9.2.31}$$

where ρ, σ are observationally accessible. For galaxies that are dominated mostly by dark matter, namely dwarf spheroidals, this can be used to place a limit on the dark matter mass and mixing angle.

Another observational quantity that would be relevant for a sterile neutrino dark matter candidate is the number of effective neutrino species, N_{eff}. The standard method of obtaining the number of neutrinos from cosmology involves measuring the number of effective relativistic species from the CMB. The sterile neutrinos we consider in this work decouple while they are still relativistic (at $\sim 10 - 15\,MeV$) as discussed in section 9.4. After sterile decoupling, all the normal standard model species continue to decay/annihilate and eventually only the active neutrinos, electrons, positrons, baryons and photons remain. Each time a species decouples, the entropy of the decoupled particles is swapped into the remaining

relativistic species via entropy conservation shown in eq 9.2.19. Because $sa^3 = constant$ the standard relation that relates the temperature of ultrarelativisitic decoupled particles to the photon temperature follows:

$$\frac{T_\nu^{active}(t)}{T_\gamma(t)} = \left(\frac{4}{11}\right)^{1/3} \quad ; \quad \frac{T_{\nu_s}(t)}{T_\gamma(t)} = \left(\frac{2}{g_d}\right)^{1/3} . \tag{9.2.32}$$

After photon reheating the expression for the relativistic energy density becomes

$$\rho_{rel} = \rho_\gamma \left(1 + \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} N_{eff,0} + \frac{\rho_{\nu_s}}{\rho_\gamma}\right) \tag{9.2.33}$$

where ρ is given by 9.2.22 and $N_{eff,0} = 3.046$ is the standard result with only the active neutrinos [378]. The CMB is formed when $T_\gamma \approx 1eV$ and if the sterile neutrinos are still relativistic at this time they may contribute to N_{eff}. During matter domination prior to photon decoupling, a relativistic sterile neutrino has energy density given by

$$\rho_{\nu_s} = \frac{g_s T_{\nu_s}^4(t)}{2\pi^2}\langle y^3\rangle . \tag{9.2.34}$$

Using $\rho_\gamma = \frac{2\pi^2}{30}T_\gamma^4$ we get that

$$\frac{\rho_{\nu_s}}{\rho_\gamma} = g_s \frac{30}{4\pi^4}\left(\frac{T_{\nu_s}(t)}{T_\gamma(t)}\right)^4 \langle y^3\rangle . \tag{9.2.35}$$

So writing

$$\rho_{rel} = \rho_\gamma \left(1 + \frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\left(N_{eff,0} + \Delta N_{eff}\right)\right) \tag{9.2.36}$$

leads to the defintion

$$\Delta N_{eff} = \left(\frac{11}{4}\frac{2}{g_d}\right)^{4/3} \frac{60 g_{\nu_s}}{7\pi^4}\langle y^3\rangle \tag{9.2.37}$$

where $N_{eff} = N_{eff,0} + \Delta N_{eff}$ has been most recently measured by the Planck satellite [63]. The results above are general and all that remains is to obtain $f_d(y)$ for a particular mechanism.

9.3 QUANTUM KINETIC EQUATION

It is generally accepted that in the early universe, where temperatures and densities are larger than the QCD scale ($\sim 155MeV$), quarks and gluons are asymptotically free forming a quark-gluon plasma. As the universe expands and cools, quarks and gluons undergo two phase transitions: deconfinement/confinement and chiral symmetry breaking. Confinement and hadronization result predominantly in the formation of baryons and mesons, the lightest of which - the pions - are dominant and are the pseudo Goldsone bosons associated with chiral symmetry breaking [384]. A recent lattice QCD calculation [369] suggests that this phase transition is not first order but a rapid crossover near a critical temperature $T_{QCD} \approx 155MeV$. Pions thermalize in the plasma via strong, electromagnetic and weak interactions and are in local thermodynamic equilibrium. Their decay into leptons and active neutrinos is balanced by the inverse process as the leptons and active neutrinos are also in LTE. However if the pions (slowly) decay into sterile neutrinos, detailed balance will not be maintained as the latter are not expected to be in LTE.

As the pion is the lowest lying bound state of QCD, it is a reasonable assumption that during the QCD phase transition pions will be the most dominantly produced bound state. During this time, pions will remain in LTE with the *active* neutrinos by detailed balance $\pi \rightleftharpoons l\nu_a$. We focus on sterile neutrino ν_s production from $\pi \to l\nu_s$ which is suppressed by $|U_{ls}|^2 \ll 1$ with respect to the active neutrinos and does not maintain detailed balance. We also restrict the analysis to a scenario with no lepton asymmetry which sets the chemical potential of pions and leptons to zero. The interaction Hamiltonian responsible for this decay is

$$H_i = \sum_{l=e,\mu} \sqrt{2} G_F V_{ud} f_\pi \int d^3x \left[\bar{\Psi}_{\nu_l}(x,t) \gamma^\sigma \mathbb{L} \Psi_l(x,t) J^\pi_\sigma(\vec{x},t) + H.C. \right] \tag{9.3.1}$$

where $J^\pi_\sigma = i\partial_\sigma \pi(x,t)$ is the pseudoscalar pion current.

The buildup of the daughter particles can be described via a quantum kinetic equation

that takes the usual form of

$$\frac{dn}{dt}(q,t) = \delta n_{Gain} - \delta n_{Loss} = \mathbb{P}[n(t)] \tag{9.3.2}$$

where the gain and loss terms are obtained from the appropriate transition probabilities $|\mathcal{M}_{fi}|^2$. For this Hamiltonian, the processes relevant for neutrino build up are displayed in fig 45.

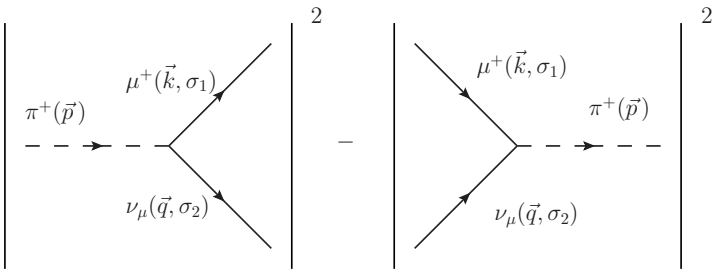


Figure 45: The gain/loss terms for the quantum kinetic equation describing $\pi^+ \to \bar{\mu}\nu_\mu$.

The gain terms arise from the reaction $\pi^+ \to \bar{l}\nu_l$ where the initial state has $N_{\vec{p}}$ quanta of pions and $n_{\vec{k},s'}, n_{\vec{q},s}$ quanta of charged leptons and neutrinos respectively while the final state has quanta $N_{\vec{p}} - 1, n_{\vec{k},s'} + 1, n_{\vec{q},s} + 1$ for the respective species. The Fock states for the gain process are given by

$$|i\rangle = |N_p^{\pi^+}, n_k^{\bar{l}}, n_q^{\nu}\rangle \quad ; \quad |f\rangle = |N_p^{\pi^+} - 1, n_k^{\bar{l}} + 1, n_q^{\nu} + 1\rangle. \tag{9.3.3}$$

Similarly, the loss terms are obtained from the reverse reaction $\bar{l}\nu_l \to \pi^+$ where the initial state has $N_{\vec{p}}, n_{\vec{k},s'}, n_{\vec{q},s}$ quanta for pions, charged leptons and neutrinos respectively. The final state has $N_{\vec{p}} + 1, n_{\vec{k},s'} - 1, n_{\vec{q},s} - 1$ of the appropriate quanta and the Fock states for the loss process are given by

$$|i\rangle = |N_p^{\pi^+}, n_k^{\bar{l}}, n_q^{\nu}\rangle \quad ; \quad |f\rangle = |N_p^{\pi^+} + 1, n_k^{\bar{l}} - 1, n_q^{\nu} - 1\rangle. \tag{9.3.4}$$

The neutrino flavor states are expanded in terms of the mass eigenstates via the U_{PMNS}

matrix as per usual

$$|\nu_\alpha\rangle = \sum_i U^*_{\alpha i}|\nu_i\rangle \tag{9.3.5}$$

Through a standard textbook calculation, the transition amplitudes at first order in perturbation theory can be calculated. The matrix element relevant for the gain term is given by

$$\begin{aligned}
\mathcal{M}_{fi}|_{gain} &= -i\int d^4x\langle N_p^{\pi^+}-1, n_k^{\bar{l}}+1, n_q^{\nu}+1|\mathcal{H}_I(x)|N_p^{\pi^+}, n_k^{\bar{l}}, n_q^{\nu}\rangle \\
&= i\sqrt{2}G_F V_{ud} f_\pi \sum_i U^*_{li}\frac{2\pi}{\sqrt{V}}\frac{\bar{\mathcal{U}}^{\nu_i}(q,\sigma_1)\not{p}\mathbb{L}\mathcal{V}^l(k,\sigma_2)}{\sqrt{8E_\pi(p)E_l(k)E_\nu(q)}} \\
&* \quad \delta_{\vec{p},\vec{k}+\vec{q}}\delta(E_\pi(p)-E_\nu(q)-E_l(k)\sqrt{N_\pi(p)}\sqrt{1-n_l(k)}\sqrt{1-n_\nu(q)}
\end{aligned} \tag{9.3.6}$$

and the matrix element relevant for the loss term is given by

$$\begin{aligned}
\mathcal{M}_{fi}|_{loss} &= -i\int d^4x\langle N_p^{\pi^+}+1, n_k^{\bar{l}}-1, n_q^{\nu}-1|\mathcal{H}_I(x)|N_p^{\pi^+}, n_k^{\bar{l}}, n_q^{\nu}\rangle \\
&= i\sqrt{2}G_F V_{ud} f_\pi \sum_i U_{li}\frac{2\pi}{\sqrt{V}}\frac{\bar{\mathcal{V}}^l(k,\sigma_2)\not{p}\mathbb{L}\mathcal{U}^{\nu_i}(q,\sigma_1)}{\sqrt{8E_\pi(p)E_l(k)E_\nu(q)}} \\
&* \quad \delta_{\vec{p},\vec{k}+\vec{q}}\delta(E_\pi(p)-E_\nu(q)-E_l(k))\sqrt{N_\pi(p)+1}\sqrt{n_l(k)}\sqrt{n_\nu(q)}
\end{aligned} \tag{9.3.7}$$

Restricting our attention towards the production of one particular mass eigenstate, $i = s$ (ie $\pi \to \bar{\mu}\nu_i$ as opposed to $\pi \to \bar{\mu}\nu_\mu$) will give the production distribution of a sterile neutrino. The idea is that the active neutrinos will remain in thermal and chemical equilibrium through $\pi \rightleftarrows \bar{l}\nu_l$ but if we assume that there had been *no sterile neutrino production* prior to pion decays then this will be the dominant contribution to sterile neutrino population. With this adjustment, summing over $\vec{k}, \vec{p}, \sigma_1, \sigma_2$ leads to the production rate

$$\begin{aligned}\frac{1}{T}\sum_{\vec{k},\vec{p},\sigma_1,\sigma_2}|\mathcal{M}_{fi}|^2_{gain} &= |U_{ls}|^2|V_{ud}|^2\frac{\pi G_F^2 f_\pi^2}{2}\int\frac{d^3p}{(2\pi)^3}\frac{N_\pi(p)(1-n_l(k))(1-n_\nu(q))}{E_\pi(p)E_\mu(k)E_l(q)} \\ &* \quad Tr[\not{p}\mathbb{L}(\not{q}+m_s)\not{p}\mathbb{L}(\not{k}-m_l)]\delta_{\vec{p},\vec{k}+\vec{q}}\delta(E_\pi(p)-E_\nu(q)-E_l(k)) \\ &= \frac{|U_{ls}|^2|V_{ud}|^2G_F^2f_\pi^2}{8\pi^2}\int\frac{d^3p}{(2\pi)^3}\frac{N_\pi(p)(1-n_l(k))(1-n_\nu(q))}{E_\pi(p)E_l(k)E_\nu(q)} \\ &* \quad [2(p\cdot q)(p\cdot k)-p^2(q\cdot k)]\delta_{\vec{p},\vec{k}+\vec{q}}\delta(E_\pi(p)-E_\nu(q)-E_l(k))\end{aligned} \tag{9.3.8}$$

where T stands for the total interaction time, not to be confused with temperature, and the evaluation of the matrix elements is a standard exercise. The loss term is calculated in the same way but with the substitution $N \to 1+N$ and $1-n \to n$. With the aforementioned replacements and using the energy/momentum conserving delta functions leads to the rate equation

$$\begin{aligned}\frac{dn}{dt} &= \frac{1}{T}\sum_{\vec{k},\vec{p},\sigma_1,\sigma_2}|\mathcal{M}_{fi}|^2_{gain}-|\mathcal{M}_{fi}|^2_{loss} \\ &= \frac{|U_{ls}|^2|V_{ud}|^2G_F^2f_\pi^2}{8\pi}\frac{m_\pi^2(m_l^2+m_\nu^2)-(m_l^2-m_\nu^2)^2}{qE_\nu(q)} \\ &* \quad \int_{p_-}^{p_+}\frac{dp\,p}{\sqrt{p^2+m_\pi^2}}\Big[N_\pi(p)(1-n_{\bar{l}}(\vec{p}-\vec{q}))(1-n_\nu(q))-(1+N_\pi(p))n_{\bar{l}}(\vec{p}-\vec{q})n_\nu(q)\Big]\end{aligned} \tag{9.3.9}$$

where $p_\pm$ are obtained from the constraint

$$[(|\vec{p}|-|\vec{q}|)^2+m_l^2]^{1/2} \le E_\pi(p)-E_\nu(q) \le [(|\vec{p}|+|\vec{q}|)^2+m_l^2]^{1/2}. \tag{9.3.10}$$

This gives the solutions

$$p_\pm = \left|\frac{E_\nu(q)}{2m_\nu^2}[(m_\pi^2-m_l^2+m_\nu^2)^2-4m_\pi^2m_\nu^2]^{1/2} \pm \frac{q(m_\pi^2-m_l^2+m_\nu^2)}{2m_\nu^2}\right|. \tag{9.3.11}$$

Note that these bounds coalesce at when $m_\pi^2-m_l^2+m_\nu^2 = 2m_\pi^2m_\nu^2$ and the rate, Eq 9.3.9, vanishes simply because this corresponds to the reaction's kinematic threshold. These results are extended to the early universe by replacing the momentum with the physical momentum, $q \to Q_f = q_c/a(t)$, and use of the results from section 9.2.

9.4 NON-THERMAL STERILE NEUTRINO DISTRIBUTION FUNCTION

A body of work [373, 372, 377, 374, 376] has established that, when π's are present in the medium in LTE, the π decay constant, f_π, and π mass vary with temperature for $T \lesssim T_{QCD}$ where T_{QCD} is the critical temperature for the QCD phase transition. We account for these effects and make several simplifications by implementing the following:

- The finite-temperature pion decay constant has been obtained in both non-linear sigma models [373] and Chiral perturbation theory [372, 377, 374, 376] with the result given as

$$f_\pi^2 \to f_\pi^2(t) = f_\pi^2(0)\left(1 - \frac{T(t)^2}{6 f_\pi(0)^2}\right) \quad ; \quad f_\pi(0) = 93 MeV\,. \tag{9.4.1}$$

  This result is required in the quantum kinetic equation since production begins near $T_{QCD} \sim 155 MeV$ and continues until the distribution function freezes out. We assume prior to T_{QCD} that there are no pions and that hadronization happens instantaneously at $T \sim T_c \sim 155 MeV$.
- The mass of the pion varies with temperature as described in detail in ref [372, 377]. The finite temperature corrections to the pion mass is calculated with electromagnetic corrections in chiral perturbation theory and its variation with temperature is shown in figure 2 of [372]. In these references it can be seen that between 50 and 150 MeV the pion mass only varies between 140 and 144 MeV. Since this change is so small, we neglect the temperature variation in the pion mass and simply use its average value: $m_\pi(T) = 142 MeV$ (see fig in ref [372]).
- We assume that the lepton asymmetry in the early universe is very small so that we may neglect the chemical potential in the distribution function of the pions and charged leptons. This asymmetry is required for the Shi-Fuller mechanism but will not be present in these calculations. We will show a similar enhancement at low moment to SF but the enhancement found here is with zero lepton asymmetry.
- With the assumption that there is no lepton asymmetry, the contribution to thermodynamic quantities from $\pi^- \to l\bar{\nu}$ will be equal to that of $\pi^+ \to \bar{l}\nu$. In which case, the degrees of freedom will be set at $g_\nu = 2$ accounting for both equal particle and antipar-

ticle contributions in the case of Dirac fermions and the two different sources ($\pi^{\pm}$) for Majorana fermions. The different helicities have already been accounted by summing over spins in the evaluation of the matrix elements of the previous section.

- We assume that there had been no production of sterile neutrinos prior to the hadronization period from any other mechanisms (such as scalar decays or DW). This allows us to set the initial distribution of the sterile neutrinos to zero in the kinetic equation which implies that our results for the distribution function will be a *lower bound for the distribution function*. Any other prior sources could only enhance the population of sterile neutrinos. By neglecting the initial population, we can neglect the Pauli blocking factor of the ν's in the production term and we can also neglect the loss term (see discussion below).

After the QCD phase transition, there is an abundance of pions present in the plasma in thermal/chemical equilibrium. The pions will decay, predominantly via $\pi^{\pm} \to l^{\pm}\nu_s\,(\bar{\nu}_s)$, producing sterile neutrinos which, assuming that sterile neutrinos had not been produced up to this point, will have a negligible distribution function. The reverse reaction ($\bar{l}\nu_s \to \pi$) will not occur in any significant quantities also due to the assumption of null initial population and $|U_{ls}|^2 \ll 1$; under these assumptions we may neglect the loss terms in the kinetic equation. With these assumptions, we use the following distributions for the production terms in the quantum kinetic equation

$$N_\pi = \frac{1}{e^{E_\pi(p,t)/t} - 1} \;\; ; \;\; n_l = \frac{1}{e^{E_l(p,t)/t} + 1} \;\; ; \;\; n_{\nu_s} \approx 0 \;\; ; \;\; E_\alpha(k,t) = \sqrt{\frac{k_c^2}{a(t)^2} + m_\alpha^2} \tag{9.4.2}$$

where k_c is a comoving momentum as discussed in section 9.2.

With these replacements, neglecting the loss terms and setting $E_l(p,q) = E_\pi(p) - E_\nu(q)$ the quantum kinetic equation becomes

$$\begin{aligned} \frac{dn}{dt}(q,t) &= \frac{|U_{ls}|^2 f_\pi(t)^2}{16\pi} \frac{m_\pi^2(m_l^2 + m_\nu^2) - (m_l^2 - m_\nu^2)^2}{q\sqrt{q^2 + m_\nu^2}} \\ &* \int_{p_-}^{p_+} \frac{dp\, p}{\sqrt{p^2 + m_\pi^2}} \left[\frac{e^{-E_\nu(q)/T} e^{E_\pi(p)/T}}{(e^{E_\pi(p)/T} - 1)(e^{-E_\nu(q)/T} e^{E_\pi(p)/T} + 1)} \right] \end{aligned} \tag{9.4.3}$$

where the limits of integration are given by Eq. 9.3.11 and we have suppressed the dependence of physical momentum on time. The integral can be done by a simple substitution with the final result given here

$$\begin{aligned}\frac{dn}{dt}(q,t) &= \frac{|U_{ls}|^2 f_\pi^2(t)}{16\pi}\frac{m_\pi^2(m_l^2+m_\nu^2)-(m_l^2-m_\nu^2)^2}{q(t)E_\nu(q,t)(e^{E_\nu(q,t)/T(t)}+1)}T(t) \\ &* \ln\left(\frac{1-e^{-\sqrt{p^2+m_\pi^2}/T(t)}}{e^{-E_\nu(q,t)/T(t)}+e^{-\sqrt{p^2+m_\pi^2}/T(t)}}\right)\Bigg|_{p=p_-(t)}^{p=p_+(t)}\end{aligned} \tag{9.4.4}$$

where $p^\pm$ are given by Eq. 9.3.11.

We make the following change of variables

$$\tau = \frac{m_\pi}{T(t)} \quad ; \quad \frac{d\tau}{dt} = \tau H(t) \quad ; \quad y = \frac{p(t)}{T(t)} = \frac{p_c}{T_0} \tag{9.4.5}$$

where T_0 is the temperature of the plasma *today* since the normalization is set by $a(t_0)=1$. The QCD phase transition begins deep inside the radiation dominated epoch as does freeze out (see below) so that the Hubble factor is given by eq 9.2.15. Inserting the form of the Hubble factor into eq 9.4.4 prompts the convenient definition

$$\Lambda = \frac{|U_{ls}|^2}{\sqrt{g(t)}}\left[\frac{|V_{ud}|^2 f_\pi^2(0) G_F^2}{8\pi * 1.66}\frac{M_{pl}}{m_\pi}\right]\left(m_l^2+m_{\nu_s}^2-\frac{(m_l^2-m_{\nu_s}^2)^2}{m_\pi^2}\right). \tag{9.4.6}$$

During the period shortly after hadronization when $m_\mu \lesssim T \lesssim m_\pi$ the relativistic degrees of freedom are $g(t)\sim 14.25$ while in the regime $m_e \lesssim T \lesssim m_\mu$ the degrees of freedom count is $g(t)\sim 10.75$ [334]. We expect the sterile neutrino decoupling to happen well above the electron mass (this will be justified later) and since the variation of $g(t)$ is small we replace it with its average value, $g(t)\sim \bar{g}=12.5$, so that we can neglect the time dependence of Λ.

These substitutions and variable changes lead to a more tractable form of the kinetic equation

$$\frac{1}{\Lambda}\frac{dn}{d\tau}(y,\tau) = \frac{(\tau/y)^2(1-\frac{m_\pi^2/6f_\pi^2}{\tau^2})}{\sqrt{1+\frac{m_{\nu_s}^2}{m_\pi^2}\frac{\tau^2}{y^2}}\left(e^{E_\nu^q/T}+1\right)}\ln\left(\frac{1-e^{-\sqrt{p^2+m_\pi^2}/T(t)}}{e^{-E_\nu(q,t)/T(t)}+e^{-\sqrt{p^2+m_\pi^2}/T(t)}}\right)\Bigg|_{p=p_-(t)}^{p=p_+(t)}. \tag{9.4.7}$$

The population build up is obtained by integrating

$$n(\tau, y) = \int_{\tau_0}^{\tau} d\tau' \frac{dn}{d\tau}(\tau', y)\,. \tag{9.4.8}$$

where we have neglected any early population of ν_s and the value of τ_0 is determined by when the pions are produced, assumed almost immediately after the hadronization transition. Our assumption is that this happens instantaneous at the QCD phase transition and the pions reach equilibrium instantaneously. This is justified by the results of [369] which suggest a continuous transition which allows for thermalization on strong interaction time scales.

As shown in [369], the continuous phase transition occurs at $T_{QCD} \sim 155MeV$ so that $\tau_0 = m_\pi/T_{QCD} = 0.92 \approx 1$. As we set τ_0 below this value we expect that the population would increase as there will simply be more time for production to occur; this will be confirmed in a subsequent section.

The rate equation and the resulting population buildup as a function of τ is shown in figures 46-48 for several values of y and m_{ν_s} for both $\pi \to \mu\nu_s$ and $\pi \to e\nu_s$. Note that the rate is enhanced for small values of y and is highly suppressed for large values of y. Fig 47 clearly illustrates that freezeout occurs by $\tau = 10$, which corresponds to temperatures $T \sim 15MeV$, for a very wide range of sterile neutrino masses.

A rough estimate of the sterile neutrino decoupling temperature can be made by considering the pion distribution. As the plasma temperature cools to well below the pion mass, the pion distribution will go as $f_\pi = e^{-m_\pi/T(t)}$ leading to a large suppression of the production rate at $T \lesssim 10MeV$ which is when we expect the sterile neutrinos to freeze out. This is indeed what is found numerically in the population build up calculations of figs 46,47,48.

9.4.1 Light mass limit

As discussed, we expect freeze out to occur on the order of $T_f \sim 10 - 15$MeV and we will consider here light sterile neutrinos with $m_{\nu_s} \lesssim \mathcal{O}(MeV)$. Restricting attention to this mass range sets $m_\nu/T_f \ll 1$ for the duration of sterile neutrino buildup and simplifies the kinetic equation considerably. For this particular production mechanism, it follows that

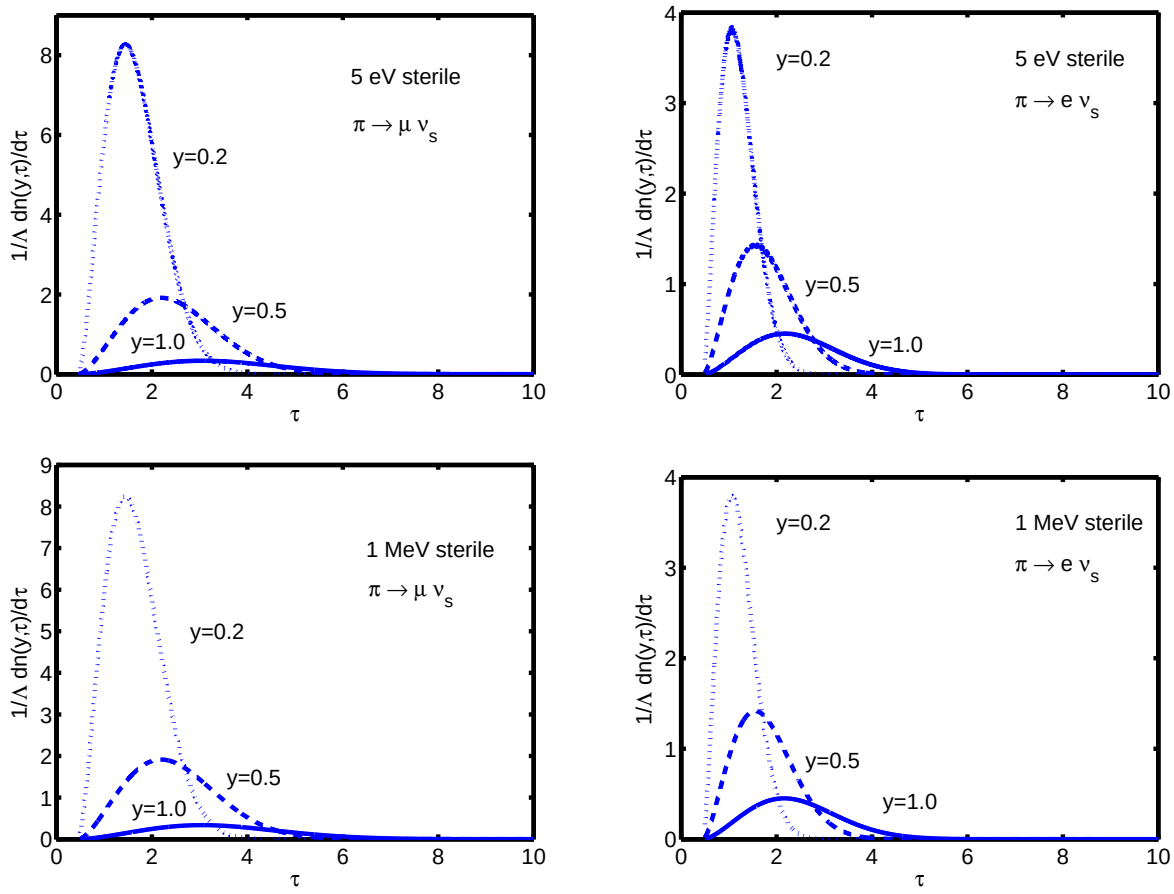


Figure 46: Production rate of a sterile ν_s obtained from quantum kinetics from $\pi \to l\nu_s$ with $l = \mu, e$. Note that for $m_\nu \lesssim 1MeV$ the rate does not vary significantly.

$m_\nu^2 \ll m_\pi^2 - m_l^2$ and we introduce the parameter

$$\Delta(m_\nu) \equiv \frac{m_\pi^2}{m_\pi^2 - m_l^2 + m_\nu^2} \tag{9.4.9}$$

so that, upon expanding in small parameters m_ν/T_f and $m_\nu^2/(m_\pi^2 - m_l^2)$ leads to the following simplifications

$$\frac{E^\pi(p^+)}{T} = \frac{1}{\Delta}\frac{m_{\pi^2}}{m_\nu^2}y \quad ; \quad \frac{E^\pi(p^-)}{T} = \Delta y + \frac{\tau^2}{4\Delta y} \tag{9.4.10}$$

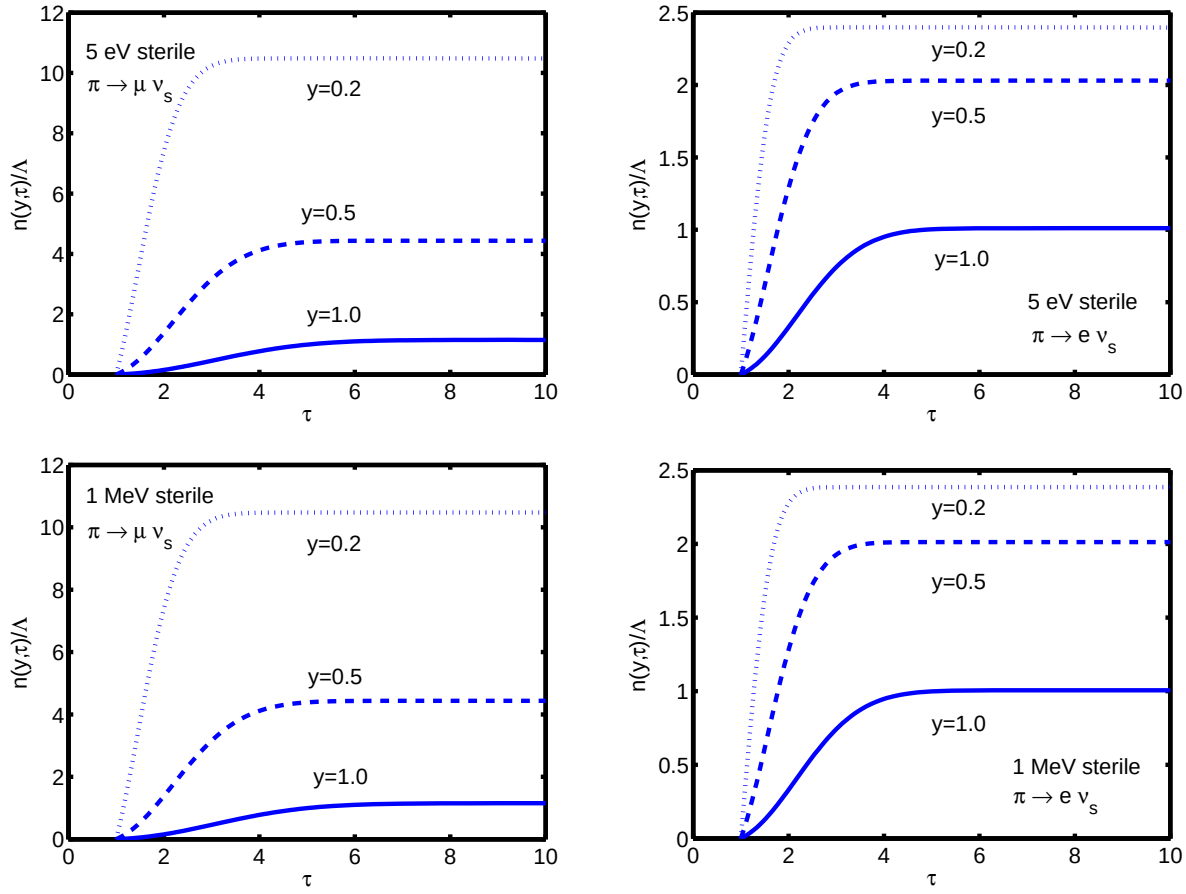


Figure 47: Population build up of a sterile ν_s obtained from quantum kinetics from $\pi \to l\nu_s$ with $l = \mu, e$. Note that for $m_\nu \lesssim 1MeV$ the build up does not vary significantly.

which is relevant for a wide range of light steriles that freeze out at $m_\nu/T_f \ll 1$. Note that we are suppressing the m_ν dependence of Δ and will do so for the remainder of this work (for $m_\nu \lesssim 1MeV$). In this limit, the kinetic equation simplifies to

$$\frac{1}{\Lambda}\frac{dn}{d\tau} = \left(\frac{\tau}{y}\right)^2 \frac{(1-\frac{m_\pi^2/6f_\pi^2}{\tau^2})}{(e^y+1)} \ln\left(\left[\frac{1-e^{-\frac{m_\pi^2}{\Delta m_\nu^2}y}}{e^{-y}+e^{-\frac{m_\pi^2}{\Delta m_\nu^2}y}}\right]\left[\frac{e^{-y}+e^{-\Delta y-\frac{\tau^2}{4\Delta y}}}{1-e^{-\Delta y-\frac{\tau^2}{4\Delta y}}}\right]\right). \tag{9.4.11}$$

We must ensure that the rate remain small in order to ignore the sterile's population build up and consequent Pauli blocking. The population scales with Λ and if one were to compute

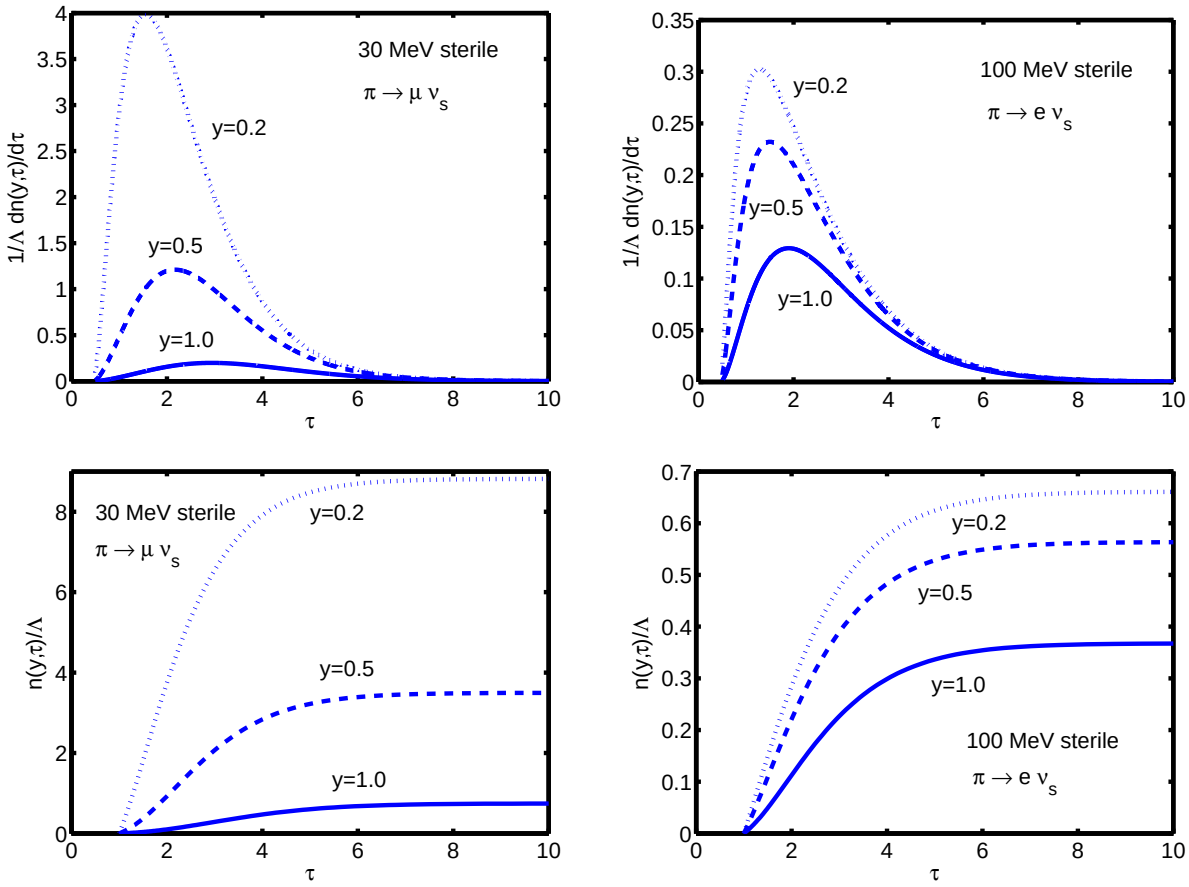


Figure 48: Rates and population build up of a sterile ν_s obtained from quantum kinetics from $\pi \to l\nu_s$ with $l = \mu, e$ and m_{ν_s} near the kinematic threshold (for μ/e production $m_{\nu_s} = 30/100 MeV$ respectively).

the next order correction by including the first order buildup in the rate equation, the higher order correction would scale as Λ^2 and so on. Provided $\Lambda \ll 1$ (discussed below), the first order correction will be sufficient and higher order perturbations will be calculated in future work.

In order to evaluate the integral analytically, several mild simplifications are made. As previously mentioned, we use the fact that $g(t)$ varies slowly during the production process and a reasonable estimate is to instead use its average value (12.5). Additionally, if we are

restricting our attention to the study of sterile neutrinos with $m_\nu \lesssim 1MeV$, then the first bracketed term inside of the logarithm (which is independent of τ) simplifies considerably.

$$\frac{1}{\Lambda}\frac{dn}{d\tau} = \left(\frac{\tau}{y}\right)^2 \frac{(1-\frac{m_\pi^2/6f_\pi^2}{\tau^2})}{(e^y+1)} \ln\left(\frac{1+e^{-(\Delta-1)y-\frac{\tau^2}{4\Delta y}}}{1-e^{-\Delta y-\frac{\tau^2}{4\Delta y}}}\right) . \tag{9.4.12}$$

The remaining τ dependence in the logarithm is a result of the Bose-Einstein suppression of the pions' thermal distribution and the $1/y^2$ dependence is a result of the phase space factors (with $m_\nu \lesssim 1MeV$).

With these simplifications and by expanding the logarithms in a power series the integral can be carried out analytically. The final result is given as

$$\begin{aligned} n(\tau,\tau_0,y) &= \frac{\Lambda}{y^2(e^y+1)}\left\{\sum_{k=1}^{\infty}\left[(-1)^{k+1}e^{-(\Delta-1)ky}+e^{-\Delta ky}\right]\right. \\ &* \left[\frac{4\Delta^{3/2}y^{3/2}}{k^{5/2}}\left(\Gamma\left(\frac{k\tau_0^2}{4\Delta y},\frac{3}{2}\right)-\Gamma\left(\frac{k\tau^2}{4\Delta y},\frac{3}{2}\right)\right)\right. \\ &- \left.\left.\frac{m_\pi^2\Delta^{1/2}y^{1/2}}{6f_\pi^2k^{3/2}}\left(\Gamma\left(\frac{k\tau_0^2}{4\Delta y},\frac{1}{2}\right)-\Gamma\left(\frac{k\tau^2}{4\Delta y},\frac{1}{2}\right)\right)\right]\right\} \end{aligned} \tag{9.4.13}$$

where $\Gamma(z,\nu)$ is the upper incomplete gamma function.

To get the frozen distribution, we take the long time limit, $\tau\to\infty$, to arrive at

$$\begin{aligned} n(\tau,\tau_0,y)\Big|_{\tau\to\infty} &= \frac{\Lambda}{y^2(e^y+1)}\left\{\sum_{k=1}^{\infty}\left[(-1)^{k+1}e^{-(\Delta-1)ky}+e^{-\Delta ky}\right]\right. \\ &* \left.\left[\frac{4\Delta^{3/2}y^{3/2}}{k^{5/2}}\Gamma\left(\frac{k\tau_0^2}{4\Delta y},\frac{3}{2}\right)-\frac{m_\pi^2\Delta^{1/2}y^{1/2}}{6f_\pi^2k^{3/2}}\Gamma\left(\frac{k\tau_0^2}{4\Delta y},\frac{1}{2}\right)\right]\right\} \end{aligned} \tag{9.4.14}$$

which can be written in a slightly different manner

$$\begin{aligned} n(\tau,\tau_0,y)\Big|_{\tau\to\infty} &= f_d(\tau_0,y) = \frac{\Lambda}{y^2(e^y+1)}\sum_{k=1}^{\infty}\left[1+(-1)^{k+1}e^{ky}\right]\frac{e^{-k\Delta y}}{k}J_k(\tau_0,y) \\ J_k(\tau_0,y) &= 2\tau_0\left(\frac{\Delta y}{k}\right)e^{-k\tau_0^2/4\Delta y}+\left(\frac{\Delta y}{k}\right)^{1/2}\left[\frac{2\Delta y}{k}-\frac{m_\pi^2}{6f_\pi^2}\right]\Gamma\left(\frac{k\tau_0^2}{4\Delta y},\frac{1}{2}\right) . \end{aligned} \tag{9.4.15}$$

Eq 9.4.15 is the decoupled distribution function of sterile neutrinos with $m_\nu \lesssim 1MeV$ arising from pion decay near the QCD phase transition. This distribution function is valid for a wide range of sterile neutrino masses as we have only assumed $m_\nu/T(t) \ll 1$ for the period of production/freezeout ($T_f \sim 10 - 15MeV$), which is valid as long as we consider $m_\nu \lesssim 1MeV$.

Note that the distribution function depends on the lower limit τ_0. The distribution function is plotted for several values of τ_0 in figure 49 where it can be seen that decreasing the lower limit increases the value of the distribution function. This is interpreted quite simply: production of steriles begins sooner and so the overall population is larger. If there are pions present in the plasma prior to the hadronization transition then this could be extended back to temperatures until the finite temperature corrections to the pion decay constant are no longer reliable: $\tau \sim m_\pi/\sqrt{6}f_\pi \sim 0.623$.

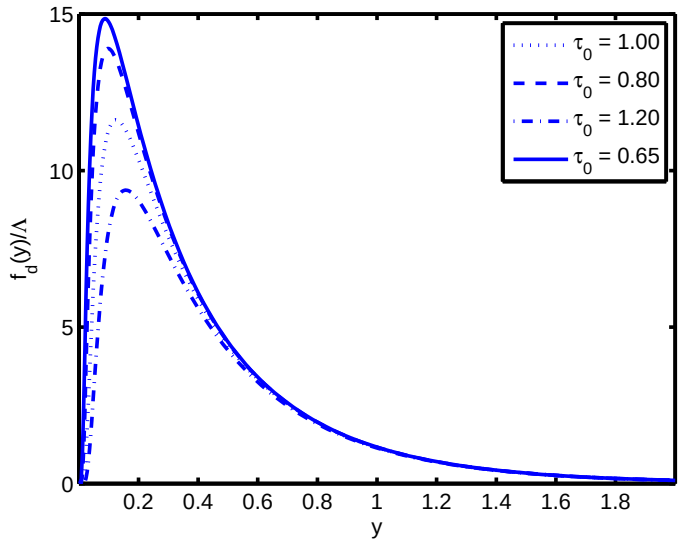


Figure 49: The distribution function with various values of initial time. Note that an earlier initial time provides an enhancement with respect to later times.

To see how this distribution differs from thermal distributions it is instructive to take the $y \to \infty$ and $y \to 0$ limits. Using that $\Gamma(k\tau_0^2/4\Delta y, 1/2) \to \Gamma(1/2) = \sqrt{\pi}$ as $y \to \infty$ gives

the asymptotic form

$$\frac{f_d(y,\tau_0)}{\Lambda}\Big|_{y\to\infty} = 2\sqrt{\pi}\Delta^{3/2}\sum_{k=q}^{\infty}\left(\frac{1+(-1)^{k+1}e^{ky}}{k^{5/2}}\right)\frac{e^{-(1+\Delta)ky}}{y^{1/2}} + \mathcal{O}\left(\frac{1}{y^{3/2}}\right)\to 0\,. \tag{9.4.16}$$

Taking the other limit $y \to 0$ along with use of the limiting expression of the Γ function, $\Gamma(x,\nu)\big|_{x\to\infty} = x^{\nu}e^{-x}$, gives the asymptotic form for $y \to 0$

$$\frac{f_d(y,\tau_0)}{\Lambda}\Bigg|_{y\to 0} = \sum_{k=1}^{\infty}\frac{\left(1+(-1)^{k+1}e^{ky}\right)}{k}\left(\frac{2\Delta}{k} - \frac{m_\pi^2}{12f_\pi^2 y}\right)\frac{\tau_0}{y}e^{-\frac{k\tau_0^2}{4\Delta y}}\to 0\,. \tag{9.4.17}$$

Both of these asymptotic forms vanish but differ widely from the asymptotic forms of thermal distributions. This serves to illustrate the highly non-thermal nature of this distribution function.

The origin of the peak in this distribution becomes clearer with these insights. At low momentum, there is a competition between the phase space factor, $1/y^2$, and the thermal pion suppression, $e^{-\tau_0^2/4\Delta y}$, which has a maximum at $y \sim \tau_0^2/4\Delta$. A low momentum enhancement occurs in the Shi-Fuller mechanism as a result of a non-zero lepton asymmetry whereas the distribution considered here features similar low momentum enhancement from a combination of thermal suppression and phase space enhancement *without the presence of a lepton asymmetry.*

Keeping the first term in the sum, $k = 1$, provides an excellent approximation to the exact result with errors of only 1%. With this approximation, the frozen distribution can be written as

$$f_d(\tau_0,y) = \frac{\Lambda}{y^2}e^{-\Delta y}\left(2\tau_0\left(\Delta y\right)e^{-\tau_0^2/4\Delta y} + \left(\Delta y\right)^{1/2}\left[2\Delta y - \frac{m_\pi^2}{6f_\pi^2}\right]\Gamma\left(\frac{\tau_0^2}{4\Delta y},\frac{1}{2}\right)\right) \tag{9.4.18}$$

where the second term is related to the error function via $\Gamma(x,1/2) = \sqrt{\pi}(1-\mathrm{erf}(\sqrt{x}))$. Note that the approximate form features a maximum for $y \simeq 1/4\Delta$ (for $\tau_0 \simeq 1$) which is confirmed numerically. This approximation is discussed below.

9.4.2 Ranges of validity

For light mass steriles, keeping just the first term in 9.4.15 is an excellent approximation. In figure 50 we have plotted both the exact distribution function and the first term of eq 9.4.15. Note that the two are nearly indistinguishable with errors only of about 1%. This approximation can be understood simply because the higher order terms in the sums ($k > 1$) feature even more exponential suppression at both small and large momentum as seen in the asymptotic expressions.

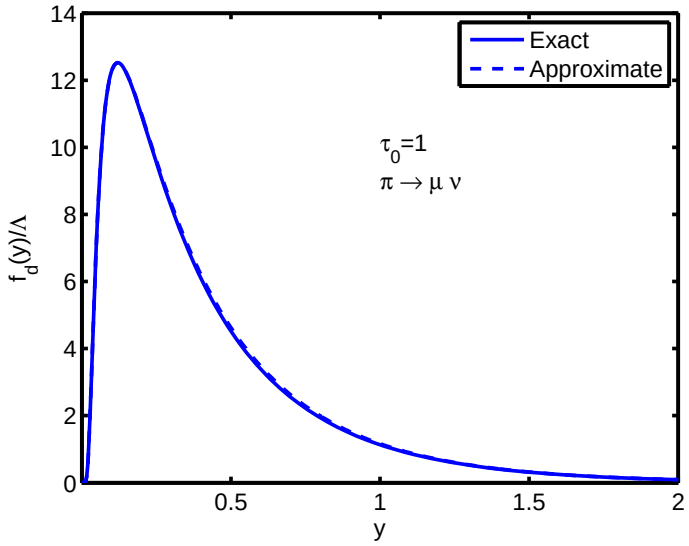


Figure 50: The exact distribution function for small mass sterile neutrinos and an approximation keeping only the first term in the series expansion.

The production process begins after $T_{QCD} \sim 155MeV$ and is complete near $T \sim 10 - 15MeV$ when the distribution freezes out. In terms of effective relativistic degrees of freedom, this implies starting with $g(T) \sim 14.25$ and concluding with $g(T) \sim 10.75$. As mentioned previously, g varies slowly which is seen in [334] so the approximation replacing $g(T)$ with its average value $\bar{g} \sim 12.5$ is reasonable.

The approximation that the neutrino population can be neglected in the quantum kinetic equation requires that $\Lambda \ll 1$. This condition arises because upon iterating the first order solution (where the population was neglected) back into the kinetic equation would result in

a perturbative expansion so that the rate equation would be of the form

$$\frac{dn}{d\tau} \sim \sum_{n=1}^{\infty} \mathcal{O}(\Lambda)^n \,. \tag{9.4.19}$$

If it were the case $\Lambda \sim \mathcal{O}(1)$ then our approximations break down and the kinetic equation would require the inclusion of higher order processes. Using the values from ref [334], the dimensionless scale Λ can be written as

$$\Lambda(T) = 6.511 \left(\frac{m_\pi^2}{GeV^2}\right) \left(\frac{|U_{ls}|^2}{10^{-5}}\right) \left(\sqrt{\frac{12.5}{g(T)}}\right) \left(\frac{m_l^2 + m_{\nu_s}^2}{m_\pi^2} - \frac{(m_l^2 - m_{\nu_s}^2)^2}{m_\pi^4}\right) \tag{9.4.20}$$

which clearly depends on the value of the sterile neutrino mass.

Taking $m_\nu \lesssim 1MeV$ implies that for $\pi \to \mu\nu$ and $\pi \to e\nu$, the neutrino mass may be neglected in the expressions for Λ. For this situation, the parameters reduce simply to

$$\Lambda_\mu \lesssim 0.03|U_{\mu s}|^2/10^{-5} \quad ; \quad \Lambda_e \lesssim 0.5|U_{es}|^2 \tag{9.4.21}$$

so that, to leading order, neglecting the sterile population is a good approximation for small mixing.

We had investigated light sterile neutrinos with $m_{\nu_s} \leq 1MeV$ and in this range the distribution function varies negligibly with m_ν. If we want to consider sterile neutrinos with $m_\nu \gtrsim 1MeV$, the approximations made previously will break down and a full numerical evaluation of the rate equation will be needed. The focus on heavy sterile neutrinos and the effect on cosmological measurements will be the study of forthcoming work where we expect nontrivial deviations from the results presented here.

9.5 OBSERVATIONAL CONSEQUENCES

9.5.1 Bounds from dark matter and dwarf spheroidals

The sterile neutrino energy density today is given by Eq 9.2.22. Note that freezeout occurs between $\tau \sim 3-5$ or $T \sim 10-15MeV$, so that, as mentioned in the previous section, the particles are relativistic at the time of decoupling. For the light sterile neutrinos we consider here, we relate the number of relativistic species at the time of sterile decoupling to the photon temperature today by the usual relation between the plasma and photon temperatures:

$$T_{plasma}(z=0) = \left(\frac{2}{g_d}\right)^{1/3} T_{\gamma,0} \lesssim 10^{-4}eV \quad ; \quad T_{\gamma,0} = 2.35 * 10^{-4}eV\,. \tag{9.5.1}$$

For $m_\nu \gtrsim 0.01eV$, we may neglect the $(y/x)^2$ term in eq 9.2.22 (ρ) and therefore the sterile neutrinos are non-relativistic *today*:

$$\rho_{\nu,0} = g_\nu m_\nu \frac{2}{g_d}\frac{T_{\gamma,0}^3}{2\pi^2}\int dy\, y^2 f_d(q_c) = m_\nu n_\nu(t_0)\,. \tag{9.5.2}$$

With this, the contribution to the density today is obtained using the distribution function calculated in section 9.4.1 and eq 9.2.13 to give

$$\frac{\Omega_{\nu_s,0}}{\Lambda} = \frac{h^2 n_\gamma}{\rho_c}\frac{g_\nu m_\nu}{2\zeta(3) g_d}\int_0^\infty dy\, y^2 \frac{f_d(y)}{\Lambda} \equiv \frac{h^2 n_\gamma}{\rho_c}\frac{g_\nu m_\nu}{2\zeta(3) g_d} I_0(m_\nu) \tag{9.5.3}$$

where

$$I_n(m_\nu) = \int_0^\infty dy\, y^{2+n}\frac{f_d(y)}{\Lambda}\,. \tag{9.5.4}$$

When $m_\nu \lesssim 1MeV$ the moments, $I_n(m_\nu)$ do not vary significantly and, for this mass range, they may be approximated by their value at $m_\nu = 0$. We work under the assumption that $m_\nu \lesssim 1MeV$ which so that we may use the limit $I_n(0)$ in subsequent calculations and the limiting values are listed in table 1.

Table 1: Table of limiting values for the function $I_n(0)$.

$I_n(0)$; $\pi \to l\nu$			
l \ n	0	1	2
e	3.756	9.675	34.300
μ	1.830	2.140	3.426

Using the results of sec 9.2, if we consider sterile masses with $m_{\nu_s} \ll m_l$ then we may neglect the sterile mass in both Δ, Λ so that we arrive at

$$\frac{\Omega_{\nu_s,0}h^2}{\Lambda} = \frac{h^2 n_\gamma}{\rho_c}\frac{g_\nu m_\nu}{2\zeta(3)g_d}I_0(0) \tag{9.5.5}$$

Considering light scalars simplifies the scales, Λ, so that the appropriate scales in the problem are

$$\Lambda_{\pi\to l\nu}(m_\nu = 0) \equiv \Lambda_l \quad ; \quad \Lambda_\mu = 0.032 * \frac{|U_{\mu s}|^2}{10^{-5}} \quad ; \quad \Lambda_e = 1.7 * 10^{-6} * \frac{|U_{es}|^2}{10^{-5}} \tag{9.5.6}$$

so that

$$m_{\nu_s}\Lambda \leq \frac{\Omega_{DM}h^2}{n_\gamma h^2/\rho_c}\left(\frac{g_d}{g_{\nu_s}}\right)\frac{2\zeta(3)}{I_0(0)} \tag{9.5.7}$$

Using the values from [334] of $n_\gamma h^2/\rho_c = 1/25.67eV$ and $\Omega_{DM}h^2 = 0.1199$ while assuming $g_{\nu_s} = 2$ and $g_d = \bar{g} = 12.5$ leads to the bounds

$$m_{\nu_s}\frac{|U_{\mu s}|^2}{10^{-5}} \leq 0.739keV \quad ; \quad m_{\nu_s}\frac{|U_{es}|^2}{10^{-5}} \leq 7242keV\,. \tag{9.5.8}$$

As discussed in sec 9.2, the dark matter phase space density decreases over the course of galactic evolution and the primordial phase space density may be used as an upper bound to obtain limits on the mass of dark matter. Using observational values for dwarf spheroidal

galaxies from ref [359] a set of bounds complementary to those from CMB measurements can be obtained. As discussed, imposing the condition $\mathcal{D}_p \geq \mathcal{D}_{nr}$ gives us the constraint

$$\mathcal{D}_p \geq \frac{1}{3^{3/2} m_{\nu_s}^4} \frac{\rho}{\sigma^3}\Big|_{today} \tag{9.5.9}$$

Assuming, as before, that the sterile neutrino mass is much smaller than the charged lepton mass renders the phase space density independent of the sterile neutrino mass. This leads to a bound on the mass given as

$$m_{\nu_s} \geq \left[\frac{1}{3^{3/2}} \frac{\rho}{\sigma^3}\Big|_{today} \mathcal{D}_p^{-1}\right]^{1/4} . \tag{9.5.10}$$

Using the results from section 9.2, the phase space density is given as

$$\mathcal{D} = \frac{g_{\nu_s} \Lambda}{2\pi^2} \frac{I_0(0)^{5/2}}{I_2(0)^{3/2}} \tag{9.5.11}$$

so that the bound becomes

$$m_{\nu_s} \Lambda^{1/4} \geq \left(\frac{2\pi^2}{3^{3/2} g_{\nu_s}} \frac{\rho}{\sigma^3}\Big|_{today} \frac{I_2(0)^{3/2}}{I_0(0)^{5/2}}\right)^{1/4} \tag{9.5.12}$$

which can serve as a complementary bound to the limits set from Ω_{DM}. Values of the phase space density today are summarized in ref [359] and using the data from the most compact dark matter haloes leads to bounds on sterile neutrino dark matter. The halo radius (r_h), velocity dispersion (σ), phase space density today and the calculated bounds are summarized in table 2 where we chose several of the most compact dwarf spheroidal galaxies (a more thorough list is available in [359]).

Taking the minimum value from this data set translates into the bounds

$$m_\nu \left(\frac{|U_{\mu s}|^2}{10^{-5}}\right)^{1/4} \geq 0.38\text{keV} \quad ; \quad m_\nu \left(\frac{|U_{es}|^2}{10^{-5}}\right)^{1/4} \geq 6.77\text{keV} . \tag{9.5.13}$$

The bounds from dwarf galaxies can be combined with the bounds from CMB measurements of Ω_{DM} to obtain allowed regions of parameter space. The two bounds are illustrated in Fig. 51 along with the parameter values reported in ref. [351] arising from the 3.5 keV x-ray signal. If sterile neutrinos are responsible for the x-ray signal then production from $\pi \to \mu\nu$ is a

Table 2: Phase space data for compact galaxies and derived bounds on sterile neutrinos arising from pion decay.

Galaxy	$\frac{r_h}{pc}$	$\frac{\sigma}{km/s}$	$\frac{\rho/\sigma^3}{(keV)^4}$	$\frac{m_\nu \Lambda_\mu^{1/4}}{keV}\Big\vert_{\min}$	$\frac{m_\nu \Lambda_e^{1/4}}{keV}\Big\vert_{\min}$
Willman 1	19	4	0.723	1.178	1.782
Segue 1	48	4	1.69	1.456	2.204
Coma-Berenices	123	4.6	0.04	0.571	0.864
Leo T	170	7.8	0.014	0.4392	0.665
Canis Venatici II	245	4.6	0.04	0.571	0.864
Draco	305	10.1	0.0036	0.3128	0.473
Fornax	1730	10.7	$2.56*10^{-4}$	0.1615	0.2445

mechanism consistent with the data within a narrow region while sterile neutrinos produced from $\pi \to e\nu$ are not.

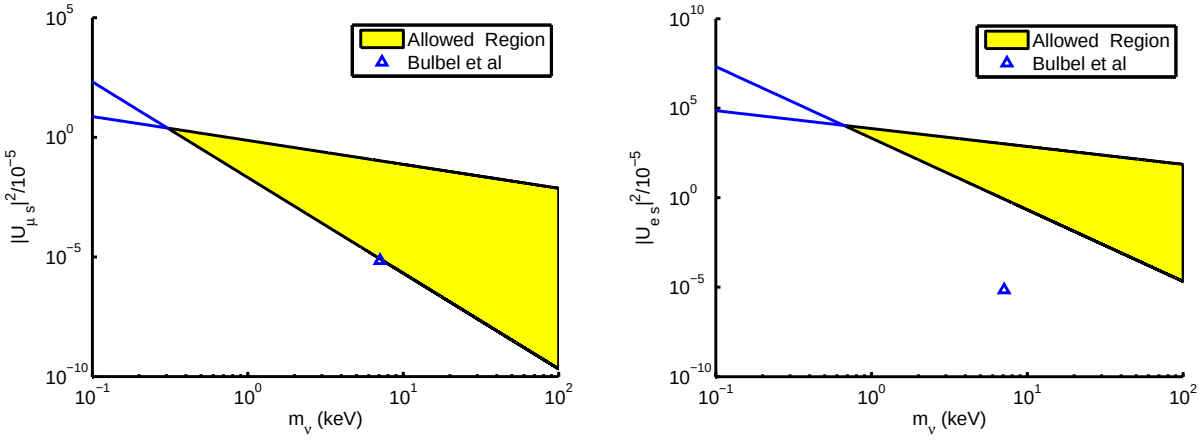


Figure 51: The bounds on sterile mass and mixing obtained from CMB and galactic measurements. The allowed regions determined from Eqs 9.5.8,9.5.13 are shaded and the sterile neutrino parameters which potentially explain the 3.5 keV signal (Bulbul et al) are also shown.

9.5.2 Equation of State and Free streaming

The equation of state for an arbitrary dark matter candidate is characterized by the parameter $w(T)$ given by eq 9.2.23. A light sterile neutrino ($m_\nu \lesssim 1MeV$) freezes out while it is still relativistic since $m/T \ll 1$ during production/freezeout therefore the results of the previous section hold. This distribution will then determine at what temperature this species becomes non relativistic via Eq 9.2.23, which is rewritten here explicitly in terms of m_ν/T:

$$w(T) = \frac{\mathcal{P}}{\rho} = \frac{1}{3}\frac{\int dy \frac{y^4}{\sqrt{y^2+\frac{m_\nu^2}{T(t)^2}}} f_d(q_c)}{\int dy\, y^2 \sqrt{y^2+\frac{m_\nu^2}{T(t)^2}} f_d(q_c)} \,. \tag{9.5.14}$$

Many fermionic dark matter candidates which freeze out at temperature T_f are treated as being in LTE in the early universe so that their distribution functions are given by the standard form

$$f_{LTE}(y) = \frac{1}{e^{\sqrt{y^2+m^2/T_f^2}}+1} \,. \tag{9.5.15}$$

To compare the new distribution to thermal results, assume that thermal particles with the same mass also freezeout while relativistic. The equation of state arising from thermal distributions and the non-thermal distribution we obtain are plotted as a function of m_ν/T in fig 52. Note that the non-thermal distribution equation of state parameter is smaller for all times. This is a reflection of the enhancement of small momentum so that the non-thermal distribution results in a dark matter species which is colder and becomes non relativistic much earlier than the thermal result. In summary, the thermal distribution produces particles that become non-relativistic when $m/T \gg 1$ whereas the pion decay mechanism produces particles that become non-relativistic when $m/T \sim 1$. This non-thermal distribution function produces a dark matter candidate that is *colder* than those produced at LTE.

The free streaming wave vector enters when one considers a linearized collisionless Bolzmann - Vlasov equation describing the evolution of gravitational perturbations which ultimately lead towards structure formation [386, 359]. The free streaming wave vector k_{fs} leads

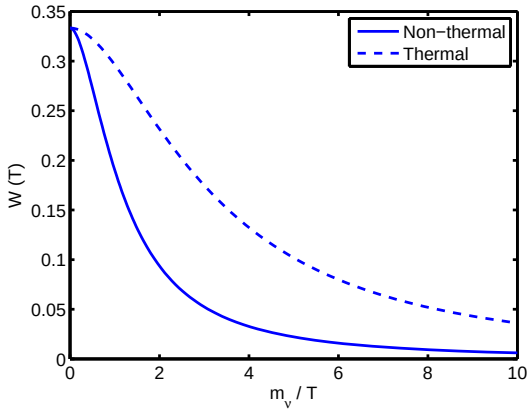


Figure 52: Equation of state compared to thermal.

to a cutoff in the linear power spectrum of density perturbations and is given by

$$k_{fs}^2 = \frac{4\pi G\rho}{\overrightarrow{V^2}}. \tag{9.5.16}$$

Modes with $k < k_{fs}$ lead to gravitation collapse in a manner akin to the Jeans instability. This is shown explicitly and discussed at length in ref [386]. Assuming that a light sterile neutrino is the only dark matter (so that $\rho_{\nu_s} = \rho_{DM}$) and using the results of section 9.2 (for a non-relativistic species), the free streaming wave vector is given by

$$k_{fs}^2 = \frac{3}{2}\frac{\Omega_{DM}H^2}{\overrightarrow{V^2}} = \frac{3}{2}H^2\Omega_{DM}\left(\frac{m_\nu}{T(t)}\right)^2\frac{\int dy\, y^2 f_d(y)}{\int dy\, y^4 f_d(y)}. \tag{9.5.17}$$

Using the latest values from Planck [63] sets the free streaming wave vector as

$$k_{fs}(z=0) = \frac{m_\nu}{T_{\gamma,0}}\left(\frac{g_d}{2}\right)^{1/3}\sqrt{\frac{3}{2}\Omega_{DM,0}H_0^2\frac{I_0(0)}{I_2(0)}} = \frac{0.617}{kpc}\frac{m_\nu}{keV}\left(\frac{g_d}{2}\right)^{1/3}\sqrt{\frac{I_0(0)}{I_2(0)}} \tag{9.5.18}$$

or in terms of the free streaming length, $\lambda_{fs} = 2\pi/k_{fs}$,

$$\lambda_{fs}(0) = 10.2kpc\left(\frac{keV}{m_\nu}\right)\left(\frac{2}{g_d}\right)^{1/3}\sqrt{\frac{I_2(0)}{I_0(0)}} \tag{9.5.19}$$

For a redshift z during matter domination the free streaming length scales as $\lambda_{fs}(z) = \lambda_{fs}(0)/\sqrt{1+z}$ and the free streaming length today for the particular processes are then given by

$$\lambda^{\mu}_{fs}(0) = 7.6\text{kpc}\left(\frac{\text{keV}}{m_\nu}\right) \quad ; \quad \lambda^{e}_{fs}(0) = 16.7\text{kpc}\left(\frac{\text{keV}}{m_\nu}\right) \tag{9.5.20}$$

where we've used the notation $\lambda^{l}_{fs}(0) \equiv \lambda_{fs}(0)\big|_{\pi\to l\nu}$.

9.5.3 Contributions to Dark Radiation

In previous sections we considered sterile neutrinos with $m_\nu \lesssim 1MeV$ specifically with $m_\nu \sim keV$ in mind. As discussed in sec 9.2, cosmological measurements can directly probe additional neutrino species through the number of effective relativistic species. We have argued that the sterile neutrinos under consideration in this work will decouple while relativistic at temperatures on the order of $10 - 15MeV$ and will remain relativistic until $T \sim m_\nu$.

In order to contribute to N_{eff}, a sterile neutrino must have mass $m_\nu \lesssim 1eV$ so that it remains relativistic through matter-radiation equality. The previous general analysis still holds but here we consider specifically sterile neutrinos with $m_\nu \lesssim 1eV$, those which are currently of interest for accelerator searches [43, 45]. The modifications to N_{eff} with the sterile neutrinos produced from pion decay are given by Eq 9.2.37 and rewritten here as

$$\Delta N_{eff} = \Lambda\frac{60g_{\nu_s}}{7\pi^4}\left(\frac{11}{2g_d}\right)^{4/3} I_1(m_\nu)\,. \tag{9.5.21}$$

As mentioned, in order to contribute to N_{eff}, the neutrinos must remain be relativistic at the time of matter-radiation equality, $T \sim eV$, so this is only valid for $m_\nu \lesssim 1eV$. In this range of masses, $I_1(m_\nu)$ does not vary appreciably and is very nearly its value for $m_\nu = 0$ which is listed in table 1. For the different processes we have

$$\Delta N_{eff}\Big|_{\pi\to\mu\nu} = 0.0040 * \frac{|U_{\mu s}|^2}{10^{-5}} \quad ; \quad \Delta N_{eff}\Big|_{\pi\to e\nu} = 9.7 * 10^{-7}\frac{|U_{es}|^2}{10^{-5}}\,. \tag{9.5.22}$$

The measurement from Planck is consistent with $\Delta N_{eff} \lesssim 0.4$ [63] and using bounds from land based experiments summarized in [343, 342] we can get an estimate of whether these light sterile will contribute significantly.

Kamland and Daya Bay [341, 337] recently reported upper bounds of $|U_{\mu s}|^2 < 0.01$ for the mass squared difference $10^{-3} eV^2 < |\Delta m_{1s}|^2 < 0.1 eV^2$. Taking the upper bound leads to $\Delta N_{eff} < 4$ suggesting that $\pi \to \mu \nu_s$ can contribute significantly to N_{eff} for a $\sim 1eV$ sterile. Ground based experiments which suggest $m_{\nu_s} \sim 1eV$ could be in tension with CMB measurements which suggest $\Delta N_{eff} \lesssim 0.4$ and $m_{\nu_s} \lesssim 0.30eV$ if the upper bound on the mixing is near its true value. Conversely, if N_{eff} could be measured more accurately, this could potentially be used to place tighter bounds on $|U_{ls}|^2$. For instance, the latest results from the Planck collaboration suggest that $\Delta N_{eff} < 0.15$ [387] which leads to the constraint $|U_{\mu s}|^2 < 3.8 * 10^{-4}$.

9.6 SUMMARY, DISCUSSION AND FURTHER QUESTIONS

We studied the production of sterile neutrinos from $\pi \to l \nu_s$ shortly after the QCD phase transition (crossover) in the early universe. Pions, being the lightest pseudoscalar mesons, are copiously produced through hadronization after the confinement-deconfinement and chiral phase transition at $T \simeq 155 MeV$ with their primary decay channel purely leptonic. Pions will be present in the plasma with a thermal distribution, maintaining LTE via strong, electromagnetic and weak interactions maintaining detailed balance (with charged leptons and active neutrinos) for kinetic and chemical equilibrium. However, pions will decay into sterile neutrinos via their mixing with active ones. We include finite temperature corrections to the pion mass and decay constant to assess the production properties of a sterile species via π decay but in absence of a lepton asymmetry.

For sterile neutrino masses $\lesssim 1 MeV$ we find that they are produced with a highly non-thermal distribution function and freeze out at $T_f \simeq 10 - 15 MeV$. The distribution function features a sharp enhancement at low momentum resulting from a competition between phase

space and thermal suppression of the parent meson. The strong low momentum enhancement featured in this non-thermal distribution function makes the species very *cold* despite such a small mass, and is remarkably similar to that found in resonant production via a lepton asymmetry [347, 355]; however, we emphasize that our study considered vanishing lepton asymmetry.

The frozen distribution function depends on a particular combination of the mass of the sterile neutrino and mixing matrix element U_{ls}. Dark matter abundance constraints from the CMB and constraints from the most dark matter dominated dwarf spheroidal galaxies provide upper and lower bounds respectively on combinations of m_s, U_{ls}. These bounds feature a region of compatibility with the recent observations of a $3.55keV$ line that could imply a $7keV$ sterile neutrino as dark matter candidate.

$$\begin{aligned} m_{\nu_s}\frac{|U_{\mu s}|^2}{10^{-5}} \leq 0.739\,\mathrm{keV} \quad &; \quad m_{\nu_s}\frac{|U_{es}|^2}{10^{-5}} \leq 7242\,\mathrm{keV} \\ m_\nu\left(\frac{|U_{\mu s}|^2}{10^{-5}}\right)^{1/4} \geq 0.38\,\mathrm{keV} \quad &; \quad m_\nu\left(\frac{|U_{es}|^2}{10^{-5}}\right)^{1/4} \geq 6.77\,\mathrm{keV} \end{aligned} \tag{9.6.1}$$

An important characteristic for structure formation is the free streaming wavevector and length, $k_{fs} = 2\pi/\lambda_{fs}$, where k_{fs} determines a cutoff in the linear power spectrum of density perturbations and consequently λ_{fs} determines the length scale below which gravitation collapse is suppressed. This scale is determined by the distribution function at freeze-out and the mass of the (non-relativistic) DM component. We find that the highly non-thermal distribution function from π decay determines that this DM species is *colder* with a $\lambda_{fs} \simeq$ few kpc today, consistent with the scale of cores observed in dwarf spheroidal galaxies. We find (today)

$$\lambda^{\mu}_{fs}(0)^2 = 7.6\,\mathrm{kpc}\left(\frac{\mathrm{keV}}{m_\nu}\right) \quad ; \quad \lambda^{e}_{fs}(0)^2 = 16.7\,\mathrm{kpc}\left(\frac{\mathrm{keV}}{m_\nu}\right) \tag{9.6.2}$$

If the mass of sterile neutrinos is $m_{\nu_s} < 1eV$ they may contribute to the radiation component between matter radiation equality and photon decoupling thereby contributing to the effective number of relativistic degrees of freedom N_{eff}. The most recent accelerator

and astrophysical bounds on the masses and mixing angles of sterile neutrinos in 3+1 or 3+2 schemes [341, 337, 343, 342] combined with the result for the frozen distribution function suggest substantial contributions from this species to N_{eff} although severe tensions remain between accelerator data and Planck bounds from the CMB

$$\Delta N_{eff}\Big|_{\pi\to\mu\nu} = 0.0040 * \frac{|U_{\mu s}|^2}{10^{-5}} \quad ; \quad \Delta N_{eff}\Big|_{\pi\to e\nu} = 9.7 * 10^{-7}\frac{|U_{es}|^2}{10^{-5}}\,. \tag{9.6.3}$$

Further Questions

While we focused on "light" sterile neutrinos with $m_{\nu_s} < 1MeV$, there are potentially important aspects to be studied for the case of $10MeV \lesssim m_{\nu_s} \lesssim 140MeV$, a range of masses kinematically available in $\pi \to e\nu_s$. These "heavier" species may actually contribute as a CDM component since freeze-out still occurs at a scale $T_f \sim 10 - 15MeV$ therefore this species will be non-relativistic and *non-thermal* upon freeze out. Heavy sterile neutrinos may decay into lighter active neutrinos on time scales larger than that for BBN. These late-produced active neutrinos would be injected into the cosmic neutrino background *after* neutrino decoupling and will therefore not be able to reach LTE with the plasma becoming a non-LTE active neutrino component which may contribute to N_{eff} non-thermally. We expect to report on these issues in further studies.

10.0 THE CASE FOR MIXED DARK MATTER FROM STERILE NEUTRINOS.

Based on: (ref. [447])

L. Lello, D. Boyanovsky, JCAP 06 (2016) 011

10.1 INTRODUCTION

The present understanding of the large scale cosmological evolution is governed by unknown quantities, dark matter and dark energy [317], where the commonly accepted wisdom points towards a dark matter candidate which is a weakly-interacting, cold thermal relic (WIMP) [318]. Dark matter candidates alternative to the WIMP paradigm attract a fair amount of attention[319, 320] due to several observational shortcomings of the standard cold dark matter cosmology (ΛCDM) at small scales. N-body simulations of cold dark matter produce dark matter profiles that generically feature cusps yet observations suggest a smooth-core profile [321, 322](core-cusp problem). The same type of N-body simulations also predict a large number of dark matter dominated satellites surrounding a typical galaxy which is inconsistent with current observations [323] (missing satellites problem). Both the missing satellites and core-cusp problem can be simultaneously resolved by allowing some fraction of the dark matter to be "warm" (WDM) [324, 325, 326, 327, 328, 329] with a massive "sterile" neutrino being one popular candidate [330, 69, 331, 70, 332] - other examples include Kaluza-Klein excitations from string compactification or axions [388, 319]. The "hotness"

or "coldness" of a dark matter candidate is discussed in terms of its free streaming length, λ_{fs}, which is the cut-off scale in the linear power spectrum of density perturbations. Cold dark matter (CDM) features small ($\lesssim$ pc) λ_{fs} that brings about cuspy profiles whereas WDM features $\lambda_{fs} \sim$ few kpc which would lead to cored profiles. Recent WDM simulations including velocity dispersion suggest the formation of cores but do not yet reliably constrain the mass of the WDM candidate in a model independent manner since the distribution function of these candidates is also an important quantity which determines the velocity dispersion and thereby the free streaming length[389].

For most treatments of sterile neutrino dark matter, a nonthermal distribution function is needed in order to evade cosmological bounds [63]. The mechanism of sterile neutrino production in the early universe through oscillations was originally studied in ref. [346] (see also the reviews [40]) and in [69, 347, 390, 448, 392] sterile neutrinos are argued to be a viable warm dark matter candidate produced out of LTE in the absence (Dodelson-Widrow, DW) or presence (Shi-Fuller, SF) of a lepton asymmetry. Models in which a standard model Higgs scalar decays into a pair of sterile neutrinos at the electroweak scale (or higher) also yield an out-of-equilibrium distribution suitable for a sterile neutrino dark matter candidate [348, 349, 350, 356]. Observations of the Andromeda galaxy with Chandra led to tight constraints on the (DW) model of sterile neutrinos[393], and more recently the observation of a 3.5 keV signal from the XMM Newton X-ray telescope has been argued to be due to a 7 keV sterile neutrino[351, 352], a position which has not gone unchallenged [353, 354, 394, 395]. The prospect of a keV sterile DM candidate continues to motivate theoretical and observational studies [356, 355, 70, 357, 358, 349, 396, 397, 392]. In all of these production mechanisms, the assumption of a vanishing initial population is implemented and, as we discuss, there are a wide array of processes which are ignored with this simplification - a problem which is starting to become appreciated [398].

On the particle physics front, neutrino masses, mixing and oscillations are the first evidence yet of physics beyond the standard model. A significant world-wide experimental program has brought measurements of most of the parameters associated with neutrino mass [334, 335] with several significant remaining questions poised to be answered in the near future [336]. Short baseline neutrino oscillation experiments (LSND, MiniBooNE)

[43, 45] present a picture of the neutrino sector which may require an additional sterile neutrino species of mass $\sim 1eV$ [332, 343, 342] but there remains tension with other experiments [337] and a definitive resolution of these anomalies will require further experiments [340, 338, 139, 89, 399, 345]. An interpretation of short baseline experiment anomalies as a signal for sterile neutrinos leads to a relatively light mass $\sim eV$ for the sterile neutrino; however, it has been argued that sterile neutrinos with mass on the order of MeV or larger [124] can decay and could also explain the short baseline anomalies or, alternatively, heavy sterile neutrinos produced through rare decay channels could also explain the anomaly [400]. Well motivated proposals make the case for a robust program to search for multiple heavy neutrinos [401, 402] in a wide range of experiments including hadron colliders [403, 404, 405, 406]. Furthermore, it has been argued that heavy sterile neutrinos in the mass range $100-500\,\mathrm{MeV}$ can decay nonthermally and enable evasion of cosmological and accelerator bounds[363].

Several current experimental programs seek direct detection of sterile neutrinos : the KATRIN (Karlsruhe Tritium Neutrino Experiment) experiment[407, 408] searches for sterile neutrinos with masses up to $\lesssim 18\,\mathrm{keV}$ in tritium beta decay, the MARE experiment[76] (Microcalorimeter Arrays for a Rhenium Experiment) explores the mass range $\lesssim 2.5\,\mathrm{keV}$ in the beta decay of Rhenium 187, the ECHo experiment (Electron Capture $^{163}\,Ho$ Experiment) [409] searches for sterile neutrinos in the mass range $\lesssim 2\,\mathrm{keV}$ in beta decay of $^{163}\,Ho$. Various recent proposals make the case for searches of *heavy* neutrinos at the Large Hadron Collider[403, 404, 405, 406] and current and future underground neutrino detectors may be able to probe dark matter candidates with $\simeq$ few MeV[410]. In extensions beyond the standard model, the possibility of a hierarchy of heavy neutrinos would be natural: there is a (very wide) hierarchy of quark masses, charged lepton masses, and as is now clear a hierarchy of light massive neutrinos.

Models of many dark matter components had been proposed recently[411], these models provide the tantalizing possibility that the decay of one heavy component can seed the production of a lighter component. As we will discuss below, this possibility also arises if there is a hierarchy of heavy neutrinos.

Motivation and Goals: As discussed above the study of sterile neutrinos with masses in a wide range, from few eV through keV up to several GeV is motivated from astrophysical observations, cosmological simulations, terrestrial accelerators experiments and the new generation of experiments that will directly search for signals of sterile neutrinos. Most theoretical extensions beyond the Standard Model that provide mechanisms to generate neutrino masses invoke one or more generations of heavy "sterile" neutrinos. While the focus on sterile neutrinos as a dark matter candidate has been on the mass range of few keV (largely motivated by the cusp-vs-core and related problems of structure formation, and more recently by the possible detection of an X-ray line at $\simeq 7\,\mathrm{keV}$), *if extensions beyond the standard model allow for a hierarchy of heavy neutrinos* these *may* yield a mixture of warm, cold (and hot) dark matter. This possibility motivates us to explore possible mechanisms of production of heavier species of *massive* neutrinos and assess in particular cases their production, freeze-out and possible consequences of non-thermal distribution functions.

Mixed dark matter described by several species of massive neutrinos with non-equilibrium distribution functions will certainly evade Lyman-α constraints[412].

Rather than proposing yet new models, the purpose of this work is twofold:

- **i):** to understand the production and freeze out of heavy sterile-like neutrinos *from standard model charged and neutral current interaction vertices* under a *minimal set of assumptions*, assessing their suitability as DM candidates. We argue that different processes with different kinematic channels are available in a wide range of temperatures and produce heavy neutrinos with different contributions to their distribution functions. The total distribution function is, therefore, a *mixture* of the contributions from the various different processes.
- **ii):** to provide an explicit example within the context of the well understood effective field theory of weak interactions of pions including finite temperature corrections. Pion decay is a production mechanism that is available shortly after the QCD crossover and produces heavy neutrinos from different kinematic channels. Therefore furnish an explicit example of the *mixed* nature of the distribution function.

The *minimal set of assumptions* are the following:

- We assume that the massless flavor neutrinos fields of the standard model are related to the fields that create and annihilate the mass eigenstates as

$$\nu_{\alpha,L}(x) = \sum_{m=1,2,3} U_{\alpha m}\, \nu_{m,L}(x) + \sum_{h=4,5\cdots} H_{\alpha h}\, \nu_{h,L}(x) \tag{10.1.1}$$

  where $m = 1, 2, 3$ refer to the light mass eigenstates that explain atmospheric and solar oscillation observations, $h = 4, 5\cdots$ refer to heavy mass eigenstates. We assume a hierarchy of masses with $M_m \ll M_h$. The heavier mass eigenstates correspond to the various putative massive neutrinos with $M_h \simeq 1\,\mathrm{eV}$ to explain the LSND/MiniBooNE anomalies, or the $M_h \simeq 50 - 80\,\mathrm{MeV}$ proposed to explain these anomalies via radiative decay[124], or $M_h \simeq 7\,\mathrm{keV}$ that could explain the X-ray line, or even $M_h \simeq$ few MeV − GeV that could be a CDM candidate. We do not endorse nor adopt any particular extension beyond the Standard Model with particular mass generation mechanisms. We only assume the relationship (10.1.1) and the existence of a hierarchy of very massive neutrinos with mass scales well separated from the atmospheric and solar ones.
- We assume that $|U_{\alpha m}| \simeq \mathcal{O}(1)$ and $|H_{\alpha h}| \ll |U_{\alpha m}|$ although this is a feature of generic see-saw type mechanisms, we do *do not invoke* a particular mechanism. We then write the charged and neutral current weak interaction vertices in the mass basis and keep *only the linear terms* in $H_{\alpha h}$. We *only* consider the production of the heavy species with M_h from these weak interaction vertices in the *mass basis*. In this manner the production of heavy neutrinos is similar to the production of standard model neutrinos via charged and neutral current vertices *if the kinematics allows for the particular process to produce a heavy neutrino mass eigenstate*.
- Our last assumption is that the light mass eigenstates (active-like) $\nu_{1,2,3}$ along with all the other standard model particles are in thermal equilibrium. We will assume that the relevant processes occurred during the radiation dominated era with $T \geq 0.1\,\mathrm{MeV}$ at which standard model active neutrinos decouple.

This minimal set of assumptions allows us to implement well understood standard methods from quantum kinetic theory to describe the production of heavy neutrinos from weak

interaction vertices and asses their distribution function when the production freezes and discuss finite temperature corrections to the production processes.

Here we do *not* consider the possibility that the masses of sterile neutrinos is a consequence of Yukawa coupling to the Standard Model Higgs, as this entails a particular model. Production of heavy neutrinos from scalar decay has been discussed in refs.[349, 350, 348, 356, 413].

As a definite example we implement this program in the particular case of production of heavy neutrinos from pion decay after the QCD transition (crossover) within the well understood effective field theory of charged pion decay including finite temperature corrections to the pion decay constant and mass. This production channel is one of the most ubiquitous sources of neutrinos in accelerator experiments. Furthermore, pion decay shortly after the hadronization transition (at $t \simeq 10\,\mu s$) in the Early Universe certainly leads to neutrino production just as in accelerator experiments.

It is shown that this production mechanism yields heavy neutrinos from different channels (μ, e) which freeze out with highly non-thermal distributions, and furnishes a definite example of "kinematic entanglement" with important cosmological consequences on their clustering properties.

Brief Summary of Results:

- In the mass basis the same standard model processes that lead to the production of active-like (light) neutrinos yield heavy ("sterile-like") neutrinos if kinematically allowed. We identify several processes that lead to the cosmological production of heavy neutrinos during the radiation dominated era in a wide range of temperatures including $\gamma\nu_m \to \nu_h$ whose inverse process describes the X-ray telltale of sterile neutrinos as well as the possibility of production from *collective excitations in the medium*. To leading order in the mixing matrix elements $H_{\alpha h}$ we obtain the quantum kinetic equations and give their exact solution in terms of gain and loss rates that obey detailed balance and can be calculated with standard model rules. We analyze the possibility of thermalization and argue that heavy neutrinos with lifetimes larger than $1/H_0$ will freeze-out with non-equilibrium distribution functions. We establish the bounds from abundance, coarsed

grained phase space density (Tremaine-Gunn) and *stability* for suitable DM candidates and discuss clustering properties of heavy neutrino species in terms of the distribution function after freeze-out.

- We generalize the concept of *mixed DM* to encompass not only different species of heavy neutrinos, but also the different distributions functions of a single species of mass M_h from different production channels. We argue that heavy neutrinos produced via standard model charged and neutral current interactions are *kinematically entangled* with leptons produced in the same reaction. This kinematic entanglement leads to different distribution functions for different channels of production of a given species, some colder than others depending on the mass of the lepton, quantified by the equation of state in each case. The total distribution function is therefore a *mixture* of all the different distribution functions from each channel. If the heavy neutrino does not thermalize (and those with lifetimes larger than $1/H_0$ do not), its distribution function at freeze-out exhibit memory of this kinematic entanglement. This memory is manifest in the equation of state, velocity dispersion, coarse grained phase space density and free streaming length.
- As a specific example, we study the production of heavy neutrinos after the QCD transition (crossover) from pion decay within the effective field theory description of weak interactions of pions including finite temperature corrections to the pion decay constant and mass in a large kinematically allowed window $M_h \lesssim 140\,\mathrm{MeV}$. Specifically we study in detail the distribution function from both available channels, $\pi \to \mu\nu_h$; $\pi \to e\nu_h$, which furnishes a clear example of the "kinematic entanglement" with the corresponding charged lepton. The distribution function from the μ channel is distinctly *colder* than that from the e channel and the total distribution function in this case is a sum (mixture) of both components, with concentrations depending on the *ratio* of mixing matrix elements $|H_{\alpha h}|^2$ for each channel. We assess heavy neutrinos produced via this mechanism as possible DM candidates by analyzing the allowed region of parameters that fulfills the abundance, phase space density and stability constraints, highlighting how the phenomenon of kinematic entanglement is manifest along the boundaries of the allowed regions. We also study the equation of state (velocity dispersion) and free streaming length (cutoff in the power spectrum) and show that both interpolate between

the colder (μ-channel) and warmer (e-channel) components as a function of M_h and the *ratio* $|H_{eh}|^2/|H_{\mu h}|^2$ showing explicitly the *mixed* nature of the distribution function.

- We *conjecture* that heavy neutrinos with lifetimes *shorter* than $1/H_0$ produced during the radiation dominated era may decay in a *cascade* into active-like neutrinos well after Big Bang Nucleosynthesis, providing a late injection of neutrinos out of thermal equilibrium into the cosmic neutrino background, and possibly, into lighter (but still heavy) and stable(r) neutrinos that could be DM candidates after matter radiation equality.

10.2 PRODUCTION AND FREEZE-OUT: QUANTUM KINETICS

Sterile neutrinos are $SU(2)$ weak singlets that do not interact via standard model weak interactions, they only couple to the massless, flavor active neutrinos via an off diagonal mass matrix. This is the general description of sterile neutrinos, different models propose different types of (see-saw like) mass matrices. For our purposes the only relevant aspect is that the diagonalization of the mass matrix leads to mass eigenstates, and these are ultimately the relevant degrees of freedom that describe dark matter in its various possible realizations, cold, warm or hot. Cold and warm dark matter would be described by heavier species, whereas the usual (light) active-like neutrinos could be a hot dark matter component depending on the absolute value of their masses. As mentioned above in this article we focus on the production of heavy neutrinos from *standard model* couplings: in the mass basis, heavy neutrinos couple via standard model charged and neutral current interactions albeit with very small mixing matrix elements. If kinematically allowed, the same processes that produce active neutrinos will produce heavy (sterile-like) neutrinos with much smaller branching ratios.

We consider the standard model augmented by neutrino masses under the assumption of a hierarchy of heavy neutrinos with masses much larger than those associated with the (light-active like) atmospheric and solar neutrinos. Upon diagonalization of the mass matrix, the active (massless) flavor neutrino fields are related to the neutrino fields that create/annihilate

mass eigenstates as

$$\nu_{\alpha,L}(x) = \sum_{m=1,2,3} U_{\alpha m}\, \nu_{m,L}(x) + \sum_{h=4,5\cdots} H_{\alpha h}\, \nu_{h,L}(x) \tag{10.2.1}$$

where $\alpha = e, \mu, \tau$, $m = 1, 2, 3$ refer to the light (atmospheric, solar) mass eigenstates with massess M_m and $h = 4, 5 \cdots$ refer to heavy mass eigenstates of mass M_h. The form of the mixing matrix is taken in a way that is *model independent* in that no assumptions are made as to the particular source of the neutrino masses and mixing. No specification of the number of Dirac or Majorana mass terms nor their source is made. In particular, we do not assume that any Dirac mass terms are generated through Yukawa couplings to the Standard Model Higgs as this constitutes a choice of one particular model. We focus *solely* on production processes from standard model charged and neutral interactions.

As discussed above our main assumptions are

$$|U_{\alpha m}| \simeq \mathcal{O}(1) \gg |H_{\alpha h}| \quad ; \quad M_h \gg M_m \tag{10.2.2}$$

and we assume that charged and neutral vector bosons, quarks, leptons and light $\nu_{1,2,3}$ neutrinos are all in thermal equilibrium. This assumption is justified with the usual arguments that the standard model interaction rates are much larger than the expansion rates down to $T \simeq 1\,\mathrm{MeV}$ when active-like neutrinos decouple from the plasma. To present the arguments in the simplest terms, we will consider vanishing chemical potentials for all relevant species, *a posteriori*, it is straightforward to include lepton asymmetries with chemical potentials. Furthermore, we will not distinguish between Dirac and Majorana neutrinos as the difference typically results in factors 2 in various transition probabilities. Lastly, we will not consider the possibility of CP-violation, implying that forward and backward (gain and loss) transition probabilities are the same. All of these assumptions can be relaxed with the concomitant complications which will be addressed elsewhere.

In the mass basis the full Lagrangian density is

$$\mathcal{L} = \mathcal{L}_{mSM} + \sum_{h=4,5\cdots} \mathcal{L}_{0h} + \mathcal{L}_I\,, \tag{10.2.3}$$

where $\mathcal{L}_{mSM}$ is the Standard Model Lagrangian augmented with a diagonal neutrino mass matrix for the active like neutrinos ν_m of masses M_m $(m = 1, 2, 3)$, in this Lagragian density the charged and neutral interaction vertices are written in terms of the neutrino mass eigenstates with the mixing matrix $U_{\alpha m}$, $\mathcal{L}_{0h}$ is the free field Lagrangian density for the heavier neutrinos ν_h of masses M_h $(h = 4, 5\cdots)$ and to linear order in $H_{\alpha h}$

$$\begin{aligned}\mathcal{L}_I &= -\frac{g_w}{\sqrt{2}} W^+_\mu \sum_{\alpha=e,\mu,\tau} \sum_{h=4,5\cdots} H^*_{h\alpha} \overline{\nu}_h(x) \gamma^\mu \mathcal{P}_L l_\alpha(x) + h.c \\ &- \frac{g_w}{2\cos\theta_W} Z^0_\mu \sum_{h=4,5\cdots} \sum_{m=1,2,3} \widetilde{H}^*_{hm} \overline{\nu}_h(x) \gamma^\mu \mathcal{P}_L \nu_m(x) + h.c.\,, \end{aligned} \tag{10.2.4}$$

where

$$\widetilde{H}^*_{hm} = \sum_{\alpha=e,\mu,\tau} H^*_{h\alpha} U_{\alpha m}\,, \tag{10.2.5}$$

with $\mathcal{P}_L = (1 - \gamma_5)/2$. From now onwards, we refer to heavy neutrinos instead of "sterile" neutrinos, because unlike the concept of a sterile neutrino, heavy neutrinos do interact with standard model degrees of freedom via charged and neutral current vertices. The new mass eigenstates will undoubtedly contribute to the tightly constrained "invisible width" of the Z^0 [334]; however, the contribution from the heavier neutrinos is suppressed with respect to the light active-like neutrinos by small branching ratios $Br \simeq |H_{lh}|^2/|U_{lm}|^2 \ll 1$ and the tight constraints on the number of neutrinos contributing to the width of the Z^0 may be evaded by very small matrix elements, which is, in fact, one of the underlying assumptions of this work.

The strategy is to pass to the Hamiltonian

$$H = H_{mSM} + H_0(\nu_h) + H_I \equiv H^{(0)} + H_I \tag{10.2.6}$$

where $H^{(0)}$ is the *total* Hamiltonian of the Standard Model in the mass basis of the light (active-like) neutrinos plus the free field Hamiltonian of the heavy neutrinos $h = 4, 5\cdots$ and H_I is the interaction Hamiltonian obtained from $\mathcal{L}_I$ in (10.2.4). We now pass to the interaction picture of $H^{(0)}$ wherein

$$H_I(t) = e^{iH^{(0)}t} H_I e^{-iH^{(0)}t}\,, \tag{10.2.7}$$

namely we will obtain the quantum kinetic equations to leading order in $g_w|H_{\alpha h}| \ll g_w$ but *in principle to all orders* in the weak interaction coupling g_w (more on this issue below). In terms of the interaction vertices (10.2.4) we can obtain the quantum kinetic equations for production of massive neutrinos to order $(|H_{\alpha h}|)^2$ from standard master (gain-loss) equations of the form

$$\frac{dn_h(q;t)}{dt} = \frac{dn_h(q;t)}{dt}\Big|_{\rm gain} - \frac{dn_h(q;t)}{dt}\Big|_{\rm loss}\,, \tag{10.2.8}$$

where $n_h(q;t)$ is the distribution function of the heavy neutrino. This is calculated for each process: the gain term describes the increase in the population $n_h(q;t)$ by the creation of a heavy state and the loss by the annihilation, the best way to understand the calculational aspects is with a few examples.

10.2.1 Setting the stage in Minkowski space-time.

We begin by describing the main aspects in **Minkowski space time** to highlight important concepts, and generalize the formulation to an expanding cosmology in the next subsection.

To begin with consider a temperature $M_{W,Z} \lesssim T \ll T_{EW}$ where $T_{EW} \simeq 160$GeV is the temperature of the electroweak transition. In this temperature range the massive vector bosons (described as three physical degrees of freedom in unitary gauge) are populated in thermal equilibrium in the plasma with the Bose-Einstein distribution functions. If the mass of the heavy neutrino is $M_h + m_\alpha < M_{W,Z}$ the massive vector Bosons can *decay* into the massive neutrino, thereby contributing to the gain term. For example, take $W^+ \to l^+_\alpha \nu_h$; each l_α and each ν_h constitute different decay channels which lead to a *gain* contribution whereas the *loss* contribution is the inverse process $l^+_\alpha \nu_h \to W^+$. The gain and loss terms are calculated by obtaining the total transition probability per unit time to a particular channel where an explicit calculation is detailed in appendix (10.7.1). For $W^\pm \rightleftharpoons l^\pm_\alpha \nu_h$, it is

straightforward to find

$$\frac{dn_h(q;t)}{dt}\Big|_{\rm gain} = \frac{2\pi}{2E_h(q)}\int \frac{d^3k\,\overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^3 2E_W(k)2E_\alpha(p)}\, n_W(k)(1-n_\alpha(p))(1-n_h(q;t)) \times \delta(E_W(k)-E_\alpha(p)-E_h(q)) \tag{10.2.9}$$

$$\frac{dn_h(q;t)}{dt}\Big|_{\rm loss} = \frac{2\pi}{2E_h(q)}\int \frac{d^3k\,\overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^3 2E_W(k)2E_\alpha(p)}\, (1+n_W(k))n_\alpha(p)n_h(q;t) \times \delta(E_W(k)-E_\alpha(p)-E_h(q))\,, \tag{10.2.10}$$

where $p=|\vec{k}-\vec{q}|$ and

$$n_W(k)=\frac{1}{e^{E_W(k)/T}-1} \;\;;\;\; n_\alpha(p)=\frac{1}{e^{E_\alpha(p)/T}+1}\,, \tag{10.2.11}$$

and $\overline{|\mathcal{M}_{fi}|^2}$ is the usual transition matrix element for $W^+\to l^+_\alpha\nu_h$ summed over the three polarizations of the vector Boson, summed over the helicity of the charge lepton and summed over the helicity of the heavy neutrino; obviously $\overline{|\mathcal{M}_{fi}|^2}\propto g_w^2|H_{\alpha h}|^2$. Therefore the quantum kinetic equation (10.2.8) becomes of the form

$$\frac{dn_h(q;t)}{dt} = \Gamma^{<}(q)(1-n_h(q;t)) - \Gamma^{>}(q)n_h(q;t) \tag{10.2.12}$$

where the gain and loss *rates* are

$$\Gamma^{<}(q) = \frac{2\pi}{2E_h(q)}\int \frac{d^3k\,\overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^3 2E_W(k)2E_\alpha(p)}\, n_W(k)(1-n_\alpha(p))\,\delta(E_W(k)-E_\alpha(p)-E_h(q)) \tag{10.2.13}$$

$$\Gamma^{>}(q) = \frac{2\pi}{2E_h(q)}\int \frac{d^3k\,\overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^3 2E_W(k)2E_\alpha(p)}\, (1+n_W(k))n_\alpha(p)\,\delta(E_W(k)-E_\alpha(p)-E_h(q))\,. \tag{10.2.14}$$

Because the W, l_α are in thermal equilibrium the gain and loss rates obey the *detailed balance condition*

$$\Gamma^{<}(q)\,e^{E_h(q)/T} = \Gamma^{>}(q)\,, \tag{10.2.15}$$

which can be confirmed straightforwardly from the explicit expressions (10.2.13,10.2.14) using the energy conserving delta functions and the relations

$$1+n_B(E) = e^{E/T}n_B(E) \;\;;\;\; 1-n_F(E) = e^{E/T}n_F(E) \tag{10.2.16}$$

where $n_{B,F}$ are the Bose-Einstein and Fermi-Dirac distribution functions respectively.

A similar exercise yields the quantum kinetic equation for the process $Z^0 \to \overline{\nu}_m \nu_h$ and the inverse process, it is of the same form as (10.2.12) now with

$$\Gamma^{<}(q) = \frac{2\pi}{2E_h(q)} \int \frac{d^3k\, \overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^3 2E_Z(k) 2E_m(p)}\, n_Z(k)(1-n_m(p))\, \delta(E_Z(k) - E_m(p) - E_h(q)) \tag{10.2.17}$$

$$\Gamma^{>}(q) = \frac{2\pi}{2E_h(q)} \int \frac{d^3k\, \overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^3 2E_Z(k) 2E_m(p)}\, (1+n_Z(k)) n_m(p) \delta(E_Z(k) - E_m(p) - E_h(q))\,, \tag{10.2.18}$$

where now $\overline{|\mathcal{M}_{fi}|^2} \propto g_w^2 |\widetilde{H}_{hm}|^2$ is the transition matrix element for the process $Z^0 \to \overline{\nu}_m \nu_h$ from the charged current vertex in (10.2.4). Again because the vector Boson and the active-like neutrino ν_m are in thermal equilibrium, the gain and loss rates obey the detailed balance condition (10.2.15).

As highlighted above, the interaction picture corresponds to considering the usual standard model vertices with light neutrinos ν_m to *all* orders in charged and neutral current weak interactions, as well as, in principle to all orders in strong interactions. This feature, in fact, leads to many different production channels of heavy neutrinos, provided the kinematics allows the particular channels. In thermal equilibrium the typical energy of quarks or leptons whose masses are $\ll T$ is $\langle E \rangle \simeq 3.15\, T$ so that for $M_h \lesssim T$ various higher order processes are available; for example, $\overline{f} f \to Z^0 \to \nu_m \nu_h$ and its inverse process where $f, \overline{f}$ refer to quark-antiquark or lepton-antilepton or $\nu_m \overline{\nu}_m$ and similar charged current processes involving either quarks or charged leptons. If $T \simeq M_Z, M_W$ the intermediate $Z^0, W^{\pm}$ are on-shell, with $n_Z(E), n_W(E) \simeq 1$ for $E \simeq T$ and the process is "resonantly" enhanced, the gain and loss terms are of the form

$$\begin{aligned} \frac{dn_h(q;t)}{dt}\Big|_{\rm gain} &= \frac{2\pi}{2E_h(q)} \int \frac{d^3k_1 d^3k_2\, \overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^6 2E_f(k_1) 2E_{\overline{f}}(k_2) 2E_m(p)} \qquad (10.2.19) \\ &\times\ n_f(k_1) n_{\overline{f}}(k_2)(1-n_m(p))(1-n_h(q;t))\, \delta(E_f(k_1) + E_{\overline{f}}(k_2) - E_m(p) - E_h(q)) \\ \frac{dn_h(q;t)}{dt}\Big|_{\rm loss} &= \frac{2\pi}{2E_h(q)} \int \frac{d^3k_1 d^3k_2\, \overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^6 2E_f(k_1) 2E_{\overline{f}}(k_2) 2E_m(p)} \qquad (10.2.20) \\ &\times\ (1-n_f(k_1))(1-n_{\overline{f}}(k_2)) n_m(p) n_h(q;t)\, \delta(E_f(k_1) + E_{\overline{f}}(k_2) - E_m(p) - E_h(q)) \end{aligned}$$

where $p = |\vec{k}_1 + \vec{k}_2 - \vec{q}|$ and the contribution from the intermediate vector Boson is included in $\overline{|\mathcal{M}_{fi}|^2}$, wherein the Z^0 propagator must include the width (with the corresponding thermal

contributions) because for $T \simeq M_Z$ the intermediate state goes "on shell" and is enhanced. This results in a region in the phase space integrals with a large "resonant" contribution but with width $\simeq \Gamma_Z$. A similar enhancement and treatment arises for processes of the form $\nu_m l_\alpha \to W \to l_\alpha \nu_h$ for $T \simeq M_W$ where the intermediate W becomes resonant. Obviously the difference (gain-loss) looks just like a typical Boltzmann equation, however, this is an equation for the *production* of the heavy neutrino ν_h as the distribution functions of $f, \overline{f}, \nu_m$ are all in thermal equilibrium as per our previous assumption. For this contribution it follows that $\overline{|\mathcal{M}_{fi}|^2} \propto g_w^4 |\widetilde{H}_{mh}|^2$, for temperatures $T \ll M_Z$ the Fermi limit implies that the production rate is $\propto G_F^2 T^5 \mathcal{F}_1[E_h/T] |\widetilde{H}_{mh}|^2$ and similar contributions with charged current exchange, of the form $G_F^2 T^5 \mathcal{F}_2[E_h/T] |H_{mh}|^2$ on dimensional grounds, with $\mathcal{F}_{1,2}$ dimensionless functions of their arguments. A similar quantum kinetic equation describes the (gain) process $f_1 \overline{f}_2 \to W \to l_\alpha \nu_h$ and its inverse (loss) process where f_1, f_2 correspond to either up/down-type quarks or l_β, ν_m for charged current interactions with the concomitant change in $|\mathcal{M}_{fi}|^2 \propto g_w^4 |H_{\alpha h}|^2$. It is clear that for these cases the general form of the quantum kinetic equation (10.2.12) holds where again $\Gamma^<, \Gamma^>$ obey the detailed balance condition (10.2.15).

Another process that is important and relevant in cosmology is the production of ν_h via $e^+ e^- \nu_m \to \nu_h$, the inverse process corresponds to the decay $\nu_h \to e^+ e^- \nu_m$ and a similar process mediated by neutral currents $\nu_m \nu_m \overline{\nu}_n \to \nu_h$ and the corresponding decay inverse process. In both cases the gain and loss terms are of the form

$$\begin{aligned}
\frac{dn_h(q;t)}{dt}\Big|_{\rm gain} &= \frac{2\pi}{2E_h(q)} \int \frac{d^3k_1 d^3k_2\, \overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^6 2E_1(k_1) 2E_2(k_2) 2E_3(p)}\, n_1(k_1) n_2(k_2) n_3(p) (1 - n_h(q;t)) \\
&\times \delta(E_1(k_1) + E_2(k_2) + E_3(p) - E_h(q)) \qquad (10.2.21) \\
\frac{dn_h(q;t)}{dt}\Big|_{\rm loss} &= \frac{2\pi}{2E_h(q)} \int \frac{d^3k_1 d^3k_2\, \overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^6 2E_1(k_1) 2E_2(k_2) 2E_3(p)}\, (1 - n_1(k_1))(1 - n_2(k_2))(1 - n_3(p)) \\
&\times n_h(q;t)\, \delta(E_1(k_1) + E_2(k_2) + E_3(p) - E_h(q))\,, \qquad (10.2.22)
\end{aligned}$$

where we read off the gain and loss rates

$$\begin{aligned}
\Gamma^<(q) &= \frac{2\pi}{2E_h(q)} \int \frac{d^3k_1 d^3k_2\, \overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^6 2E_1(k_1) 2E_2(k_2) 2E_3(p)}\, n_1(k_1) n_2(k_2) n_3(p) \\
&\times \delta(E_1(k_1) + E_2(k_2) + E_3(p) - E_h(q)) \qquad (10.2.23) \\
\Gamma^>(q) &= \frac{2\pi}{2E_h(q)} \int \frac{d^3k_1 d^3k_2\, \overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^6 2E_1(k_1) 2E_2(k_2) 2E_3(p)}\, (1 - n_1(k_1))(1 - n_2(k_2))(1 - n_3(p)) \\
&\times \delta(E_1(k_1) + E_2(k_2) + E_3(p) - E_h(q))\,, \qquad (10.2.24)
\end{aligned}$$

where $\vec{p} = \vec{q} - \vec{k}_1 - \vec{k}_2$ and the labels $1, 2, 3$ refer to the respective fermions either $e^{\pm}$ or ν_m for example (not to be confused with the labels for the light neutrinos ν_m). We notice that as the temperature diminishes, setting the occupation factors $n_j = 0$ in the loss term one recovers the *decay rate* of the heavy neutrino, this observation will be important in the discussion of thermalization and stability of the DM candidate below.

It is clear from this discussion that in the mass basis, standard model charged and neutral current interaction vertices lead to production processes for heavy neutrinos that are similar to those of active-like neutrinos constrained by the usual kinematics. For example at temperatures $T \gtrsim 1\,\mathrm{GeV}$, tau-lepton decay may lead to the production of heavy neutrinos with masses up to $\simeq$ GeV. Heavy lepton decay is an available mechanism down to $T \simeq M_\mu \simeq 100\mathrm{MeV}$ and are processes that have not yet been studied in detail and clearly merit attention.

Finally, at temperatures $T \gtrsim M_h$ the production (gain) process $\gamma\nu_m \to \nu_h$ is kinematically allowed, the inverse process $\nu_h \to \gamma\nu_m$ is precisely the process conjectured to yield the X-ray line as a telltale of keV neutrinos. The corresponding Feynman diagrams for the gain and loss terms are depicted in fig. (53) For these processes we find

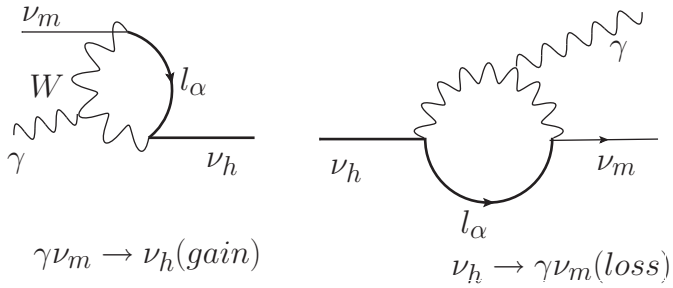


Figure 53: $\gamma\nu_m \to \nu_h$ (gain) and the inverse process $\nu_h \to \nu_m\gamma$

$$\Gamma^{<}(q) = \frac{2\pi}{2E_h(q)} \int \frac{d^3k\,\overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^3 2E_\gamma(k) 2E_m(p)}\, n_\gamma(k) n_m(p)\, \delta(E_\gamma(k) + E_m(p) - E_h(q)) \tag{10.2.25}$$

$$\Gamma^{>}(q) = \frac{2\pi}{2E_h(q)} \int \frac{d^3k\,\overline{|\mathcal{M}_{fi}|^2}}{(2\pi)^3 2E_\gamma(k) 2E_m(p)}\, (1 + n_\gamma(k))(1 - n_m(p))\delta(E_\gamma(k) + E_m(p) - E_h(q))\,, \tag{10.2.26}$$

where the $W - l_\alpha$ loop is included in the $\overline{|\mathcal{M}_{fi}|^2}$. How large can the production rate be?, at $T \ll M_W$ on dimensional grounds we expect (see discussion in section (10.2.5))

$$\Gamma^{<}_{\gamma\nu_m\to\nu_h} \propto \alpha_{em} G_F^2 T^5 \mathcal{F} \sum_l |H_{hl} U_{lm}|^2 \tag{10.2.27}$$

with $\mathcal{F}$ a dimensionless function of the ratios $M_h/T, M_l/T$ with a finite limit for $T \gg M_{h,l}$. For $T \simeq$ GeV this contribution to the production rate *could* be of the same order as that for non-resonant production (DW) at $T \simeq 150\,\text{MeV}$[69, 70, 390], clearly motivating a deeper assessment of these processes.

At $T = 0$ only the loss term (10.2.26) survives, and the corresponding decay rate has been obtained in ref.[414, 28], this will be an important aspect discussed below. The gain process is actually kinematically allowed at temperatures $T \lesssim M_h$ because of the tail in the fermionic and blackbody distributions, although suppressed at lower temperatures.

The important aspect of these latter processes is that whereas the gain rate $\Gamma^{<}(q) \to 0$ as the lepton and photon populations vanish at small temperature, the loss rate $\Gamma^{>}(q)$ *does not* vanish when the lepton populations vanish, the reason for this is that any population of ν_h *decays* into the lighter leptons. These particular processes will be important in the discussion within the cosmological context. The radiative decay $\nu_h \to \nu_m \gamma$ is conjectured to be a telltale of the presence of "sterile" (heavy) neutrinos. We then see that the inverse process *produces* the heavy (sterile-like) neutrinos at high enough temperature.

10.2.2 Finite temperature corrections:

There are important loop corrections at finite temperature, self-energy corrections to the incoming and outgoing external "legs" as well as vertex corrections. There are also finite temperature corrections to the mixing angles arising from self-energy loops, these tend to suppress the mixing matrix elements[415, 40, 416] therefore in medium the matrix elements $H_{hl} \to H_{hl,eff}(T)$ and typically in absence of MSW resonances $H_{hl,eff}(T) < H_{hl}$. An explicit example is given in section (10.4) below (see eqn. (10.4.2)). At high temperature there are hard thermal loop corrections to the self-energy of fermions and vector bosons[417, 418, 419, 420, 421] that lead to novel collective excitations with masses $\propto gT$ where g is the

gauge coupling. Photons and fermion-antifermion pairs form plasmon collective excitations with mass $\propto eT$. For $T > T_{EW}$ the W, Z vector bosons *do not* acquire a mass via the Higgs mechanism because the ensemble average of the Higgs field vanishes, but they acquire *thermal* masses $\propto g_w T$ akin to the plasmon collective excitations. Thermal masses for collective excitations of W, Z *could* open up kinematic windows for decay into ν_h *above* the electroweak transition. Plasmon collective excitations from photons in the medium can also produce ν_h via the *electromagnetic* process $\gamma^* \to \overline{\nu}_m \nu_h$ via the Feynman diagrams displayed in fig. (54). These processes are similar to the mechanism of energy loss by plasmon decay into neutrinos in highly evolved massive stars[422, 423, 424] such as red giants and are also available in the Early Universe.

Although *a priori* these processes are subleading being suppressed by higher orders in the couplings, photons are populated all throughout the thermal history of the Universe, so these processes *may* contribute to production during a longer time scale as the leading processes described above. A similar possibility *may be* associated with plasmon collective excitations for Z^0 above T_{EW} resulting in the production of the heavier species at $T > T_{EW}$, this possibility merits deeper understanding, which is beyond the realm of this article. A follow up study will be reported elsewhere.

The main point of this discussion is to highlight that in the mass basis, standard model interactions provide a wide variety of mechanisms to produce heavy neutrinos which are available at high temperatures in the early Universe. The final distribution function of a particular heavy neutrino species after freeze-out is a *mixture* of the distribution functions arising from the various production processes. A full assessment of a particular species as a DM candidate thus requires a thorough understanding of the various physical processes that lead to its production.

10.2.3 Thermalization in Minkowski space-time

An important aspect of the general form of the quantum kinetic equation (10.2.12) is the linear dependence on the (time dependent) distribution function $n_h(q;t)$. This linearity is a consequence of keeping the lowest order term in $|H_{\alpha h}|; |\widetilde{H}_{\alpha m}|$ and along with detailed balance

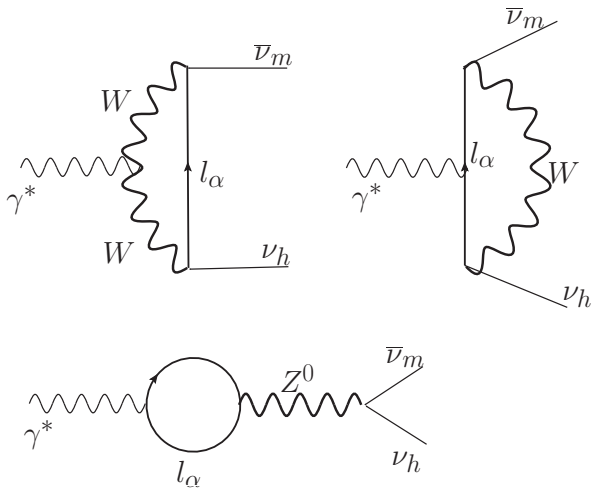


Figure 54: Gain processes: $\gamma^* \to \overline{\nu}_m \nu_h$

has profound consequences: in Minkowski space time the quantum kinetic equation (10.2.12) leads to thermal equilibration of the heavy neutrino as the following argument shows.

The solution of (10.2.12) is

$$n_h(q;t) = n_h(q;0)\, e^{-\gamma(q)t} + e^{-\gamma(q)t} \int_0^t \Gamma^<(q) e^{\gamma(q)t'}\, dt' \quad ; \quad \gamma(q) = \Gamma^<(q) + \Gamma^>(q)\,, \tag{10.2.28}$$

carrying out the time integral and using the detailed balance result (10.2.15) one finds

$$n_h(q;t) = n_{eq}(q) + \left[n_h(q;0) - n_{eq}(q)\right] e^{-\gamma(q)t}\,, \tag{10.2.29}$$

where $n_{eq}(q)$ is the equilibrium (Fermi-Dirac) distribution function. Obviously $\gamma(q)$ is the rate of relaxation towards equilibrium, for $t \gg 1/\gamma(q)$ the distribution function is that of thermal equilibrium. Since $\gamma \propto g_w^2 |H|^2$ one would be tempted to neglect the exponential terms in (10.2.28), however the exponent is *secular*, growing with time and becoming non-perturbative on a time scale $t \simeq 1/\gamma$. In Minkowski space time where the gain and loss rates are constant in time the distribution function will always reach thermal equilibrium at long time $t \gg 1/\gamma$. This is an important observation: detailed balance between the gain and loss

term guarantees that at asymptotically long time $t \gg 1/\gamma$ the heavy neutrino thermalizes and its distribution function is n_{eq}.

This is a general result *in Minkowski space time* which will be revisited below within the context of an expanding cosmology, where the relevant time scales are determined by the Hubble time scale.

10.2.4 Production, freeze-out, LTE and decay in expanding cosmology

In an expanding cosmology described by an isotropic and homogeneous Friedmann - Robertson - Walker metric, and during a radiation dominated epoch the temperature varies with time $T(t) \propto 1/a(t)$[425] where $a(t)$ is the scale factor. The energy of a particle species of mass M measured locally by an observer is

$$E(t) = \sqrt{p_f^2(t) + M^2} \quad ; \quad p_f(t) = \frac{p_c}{a(t)} \,, \tag{10.2.30}$$

where $p_f(t), p_c$ are the physical and comoving momenta respectively. The distribution function of a particle species in local thermodynamic equilibrium (LTE) is (in absence of chemical potentials)

$$n^{\pm}(E(t);t) = \frac{1}{e^{E(t)/T(t)} \pm 1} \,, \tag{10.2.31}$$

for Bose-Einstein and Fermi-Dirac respectively. The ratio

$$\frac{p_f(t)}{T(t)} = y = \frac{p_c}{T_0} \,, \tag{10.2.32}$$

is constant during a radiation dominated era and T_0 would be the temperature of the plasma today, related to the temperature of the cosmic radiation by accounting of the reheating of the photon gas whenever the number of the relativistic degrees of freedom changes[425]. As the temperature drops during the expansion when $M/T(t) \gg 1$ the population of the massive species is strongly thermally suppressed (by setting vanishing chemical potential we assume that there is no conservation of this species). In LTE the calculation of the gain and loss rates carry over from Minkowski space-time by replacing the energies and momenta by $E(t), p_f(t)$ (10.2.30) and the distribution functions of the species in LTE by (10.2.31). Although it is typical to separate the explicit time dependence in the rate equations from the

time dependence of $p_f(t)$, we will consider the rates and distribution functions as functions of p_c and the explicit time dependence of $a(t)$. Thus the quantum kinetic equation in the cosmological setting becomes

$$\frac{dn_h(q_c;t)}{dt} = \Gamma^{<}(q_c;t)(1-n_h(q_c;t)) - \Gamma^{>}(q_c;t)n_h(q_c;t)\,, \tag{10.2.33}$$

where the gain and loss rates are obtained as in Minkowski space time replacing the momenta and energies by the local time dependent counterparts in the expanding cosmology and the (LTE) distribution functions for the various bosonic or fermionic species (with vanishing chemical potentials).

As a consequence of the linearity of the quantum kinetic equation (10.2.33), which, in turn is a consequence of keeping only terms of $\mathcal{O}(|H_{\alpha h}|^2), \mathcal{O}(|\widetilde{H}_{mh}|^2)$, the full quantum kinetic equation is a simple sum over all possible channels with total gain and loss rates

$$\Gamma^{<}_{tot}(q_c;t) = \sum_{all\ \ channels} \Gamma^{<}(q_c;t) \quad ; \quad \Gamma^{>}_{tot}(q_c;t) = \sum_{all\ \ channels} \Gamma^{>}(q_c;t)\,, \tag{10.2.34}$$

therefore the gain and loss rates in the quantum kinetic equation (10.2.33) are the total rates (10.2.34).

Because for each channel the rates $\Gamma^{<,>}$ are calculated with distribution functions in (LTE) and obey the detailed balance condition this translates into

$$\Gamma^{<}_{tot}(q_c;t)\, e^{E_h(t)/T(t)} = \Gamma^{>}_{tot}(q_c;t)\,. \tag{10.2.35}$$

The general solution of (10.2.33) with the total gain and loss rates is

$$n_h(q_c;t) = n_h(q_c;t_0)\, e^{-\int_{t_0}^{t}\gamma(q_c;t')dt'} + e^{-\int_{t_0}^{t}\gamma(q_c;t')dt'} \int_{t_0}^{t} \Gamma^{<}_{tot}(q_c;t')\, e^{\int_{t_0}^{t'}\gamma(q_c;t'')dt''}\, dt'\,, \tag{10.2.36}$$

where the *relaxation rate*

$$\gamma(q_c;t) = \Gamma^{<}_{tot}(q_c;t) + \Gamma^{>}_{tot}(q_c;t) = \frac{\Gamma^{<}_{tot}(q_c;t)}{n_{eq}(t)}\,, \tag{10.2.37}$$

with

$$n_{eq}(t) = \frac{1}{e^{E_h(t)/T(t)}+1}\,, \tag{10.2.38}$$

and we used the detailed balance result (10.2.35).

A fixed point solution with $\dot{n}_h(q_c;t) = 0$ corresponds to *freeze out*. Although $\Gamma^<$ and $\Gamma^>$ are related by detailed balance, the LTE distribution function n_{eq} is not stationary, therefore is not a fixed point of the quantum kinetic equation[1]. However, the more relevant criterion for freeze-out is that the distribution function varies much more slowly than the expansion time scale, namely[379, 425],

$$\frac{\dot{n}_h(q_c;t)}{n_h(q_c;t)} \ll H(t)\,. \tag{10.2.39}$$

As discussed above, the gain rate $\Gamma^<$ depends on the population of the particles whose decay or combination results in the production of the heavy neutrino, as a result the production rate diminishes during cosmological expansion eventually vanishing exponentially because of thermal suppression of the respective LTE distribution functions. Therefore *if*

$$\int_{t_0}^{\infty} \gamma(q_c,t')\,dt' \ll 1 \tag{10.2.40}$$

we can neglect the exponential terms in the solution (10.2.36). The remaining integral is finite at asymptotically long time because the gain rate vanishes exponentially, leading to a *non-equilibrium frozen distribution*

$$n_h(q_c;\infty) \simeq n_h(q_c;t_0) + \int_{t_0}^{\infty} \Gamma^<_{tot}(q_c;t')\,dt'\,. \tag{10.2.41}$$

The freeze-out condition is achieved since (10.2.39) is fulfilled with $\dot{n}_h(q_c;t) = \Gamma^<_{tot}(q_c;t)$ vanishing exponentially at sufficiently long time, whereas $H(t) \propto 1/t$, and $n_h(q_c;\infty) \neq 0$.

When the condition (10.2.40) is valid and the distribution function is *linearly* related to the production rate as in (10.2.41) we can separate the contribution from the different production channels which will, generally, lead to *different distribution functions*, the total distribution function being a simple sum over all the different production channels. Because in each production channel the heavy neutrino is "kinematically entangled" with other leptons the distribution function at freeze-out for each channel will reflect the different kinematics, which can be interpreted as a "memory" of the particular production process. We will study a relevant example below.

[1] This is also the case for the Maxwell-Boltzmann distribution in a radiation dominated cosmology[379].

Although the gain and loss rates satisfy the detailed balance condition (10.2.35) the LTE distribution function is *not* a solution of the quantum kinetic equation (10.2.33) because $\dot{n}_{eq}(t) \neq 0$. We can assess if and when a distribution function is very nearly in LTE by exploiting the relations (10.2.35,10.2.37) to cast (10.2.33) in the form

$$\frac{dn_h(q_c;t)}{dt} = -\gamma(q_c;t)\Big[n_h(q_c;t) - n_{eq}(t)\Big]\,, \tag{10.2.42}$$

writing

$$n_h(q_c;t) = n_{eq}(t) + \delta n_h(q_c;t)\,, \tag{10.2.43}$$

and inserting into (10.2.42) one finds

$$\delta n_h(q_c;t) = \delta n_h(q_c;t_0) e^{-\int_{t_0}^{t}\gamma(q_c;t')dt'} - e^{-\int_{t_0}^{t}\gamma(q_c;t')dt'}\int_{t_0}^{t}\dot{n}_{eq}(q_c;t')\,e^{\int_{t_0}^{t'}\gamma(q_c;t'')dt''}\,dt' \tag{10.2.44}$$

with

$$\dot{n}_{eq}(t) = -n_{eq}(t)(1-n_{eq}(t))\,\frac{M_h^2\,H(t)}{T(t)E_h(t)}\,, \tag{10.2.45}$$

and

$$H(t) = 1.66\,g_{eff}^{\frac{1}{2}}(T)\,\frac{T^2(t)}{M_{Pl}} \simeq 2\times 10^5\,g_{eff}^{\frac{1}{2}}(T)\Big(\frac{T(t)}{\mathrm{GeV}}\Big)^2\,s^{-1} \tag{10.2.46}$$

during radiation domination. As the temperature drops during expansion n_{eq} diminishes rapidly and we expect that the integral with $\dot{n}_{eq}(t)$ in (10.2.44) will be finite at long time as the exponential suppression from $n_{eq}(t)$ will overwhelm the perturbative growth of the integral of γ. Hence the condition that the distribution function becomes nearly LTE, becomes

$$\int_{t_0}^{t}\gamma(q_c,t')dt' \gg 1\,. \tag{10.2.47}$$

at long time $t \lesssim 1/H_0$. If this condition is fulfilled the full solution (10.2.36) must be considered as the integrals of $\gamma(q_c,t)$ cannot be neglected. In this case the distribution function at long time may be described by (nearly) LTE, and the "memory" of the production process and the kinematic "entanglement" characteristic of each production channel is erased. Therefore, it is important to assess under what circumstances the condition (10.2.47) could be satisfied, since under such circumstances the heavy neutrino can *thermalize* with the rest of the standard model particles.

As temperature diminishes throughout the expansion history, the gain term is strongly suppressed by the thermal population factors because these processes entail the annihilation of particles in the initial state which are thermally populated (by assumption) (see the examples described above). The loss rate will also vanish exponentially by thermal suppression at long time if it involves the annihilation of *any thermal species*, therefore for these processes the condition (10.2.40) is expected to be fulfilled and the distribution function is expected to freeze out of LTE.

However for processes in which the heavy neutrino *decays* into the final products the loss term does *not* vanish at low temperatures and the distribution function eventually decays, an example of this case is the loss term (10.2.24) describing the decay processes $\nu_h \to e^+e^-\nu_m$ or $\nu_h \to \nu_{m_1}\nu_{m_2}\nu_{m_3}$. This has important implications: consider the quantum kinetic equation in the form (10.2.42) (note that $\gamma(q_c;t) > 0$) and an initial condition in the far past with $n_h(q_c;t_0) = 0$. At early times n_h *grows* $\dot{n}_h > 0$ as the gain terms dominate and the heavy neutrinos are being produced. *If* at late times the population *decays*, namely $\dot{n}_h < 0$ this means that at some time the distribution function has reached LTE at which $\dot{n}_h = 0$, as the cosmological expansion continues the loss term dominates and the population decays. This simple analysis leads to the conclusion that if the lifetime of the heavy neutrino is smaller than the age of the Universe, meaning that it is now decaying, at some point in the past its distribution function *reached LTE*. Conversely, if the lifetime is much longer than the Hubble time $1/H_0$ then the distribution function has not reached LTE and the heavy neutrino is non-thermal. In other words, a heavy neutrino that is stable during the age of the Universe $\sim 1/H_0$ and is a suitable DM candidate *must be out of LTE*.

The analysis above indicates that when the lifetime of the heavy neutrino is much smaller than the age of the Universe, its distribution function must have passed through LTE during the evolution. The LTE solution $n_{eq}(t)$ is *not* a true fixed point of the kinetic equation (10.2.33) because $\dot{n}_{eq}(E_h(t)) \neq 0$, however, we can ask what is the condition that ensures that the distribution function remains *nearly* in LTE if and when it reaches LTE. To answer this question we write

$$n_h(q_c;t) = n_{eq}(t) + \delta n_h^{(1)}(t) + \cdots \tag{10.2.48}$$

where $\delta n_h^{(1)}(t) \ll n_{eq}(t)$) etc, the expansion is in a small parameter to be quantified *a posteriori*. Introducing this ansatz into (10.2.33) we find

$$\frac{\delta n_h^{(1)}(t)}{n_{eq}(t)} = (1 - n_{eq}(t)) \, \frac{M_h^2}{T(t)E_h(t)} \, \frac{H(t)}{\gamma(q_c;t)} \,, \tag{10.2.49}$$

in terms of the relaxation time

$$\tau(q_c;t) = \frac{1}{\gamma(q_c;t)} \,, \tag{10.2.50}$$

we find that once LTE is attained, the distribution function will linger near LTE whenever

$$\tau(q_c;t)H(t) \ll \frac{T(t)}{M_h} \,. \tag{10.2.51}$$

As the temperature drops, the contribution from $\Gamma^<$ (gain) to $\gamma = \tau^{-1}$ is suppressed and the relaxation time becomes the lifetime of the heavy neutrino (loss via decay). This can be seen for example, from the loss term for the process $\nu_h \to \nu_m \nu_{m'} \nu_{m''}$ with $\Gamma^<$ given by (10.2.24) with $1, 2, 3 = m, m', m''$ and the discussion following it. Therefore the condition (10.2.51) is actually equivalent to the statement of a lifetime much shorter than the Hubble time, consistently with the discussion above. If the heavy neutrino is a *stable* DM candidate its lifetime must be $\tau > H^{-1}(t)$, implying that its distribution function will not be in LTE.

10.2.5 Stability and lifetime:

To be a suitable DM candidate a heavy neutrino must feature a lifetime $\tau \geq 1/H_0$, thus it remains to understand the decay channels for a firmer assessment of the lifetime of heavy neutrinos, their suitability as DM candidates and the conditions for non-LTE distribution functions.

The decay channels of a heavy neutrino were studied in refs.[89, 399, 28], for example the purely leptonic and radiative channels: neutral current process $\nu_h \to \nu_m \nu_{m'} \nu_{m''}$, or the charged current process $\nu_h \to e^+ e^- \nu_m$ and $\nu_h \to \nu_{h'} \gamma$; $\nu_h \to \nu_m \gamma$ these are the inverse processes associated with the production processes $\nu_m \nu_{m'} \nu_{m''} \to \nu_h$, $e^+ e^- \nu_m \to \nu_h$ and

$\nu_{h',m}\gamma \to \nu_h$ respectively. At $T=0$ the decay rates for $\nu_h \to \nu_m \nu_{m'} \nu_{m''}$; $\nu_h \to e^+ e^- \nu_m$ have been obtained in ref.[89, 399], these are given by

$$\Gamma(\nu_h \to e^+ e^- \nu_m) \simeq 3.5 \times 10^{-5} |H_{eh}|^2 \left(\frac{M_h}{MeV}\right)^5 K\left[\frac{m_e^2}{M_h^2}\right] \frac{1}{s} \tag{10.2.52}$$

where[89]

$$K\big[x\big] = (1-4x)^{1/2}(1-14x-2x^2-12x^3) + 24x^2(1-x^2)\ln\left[\frac{1+(1-4x)^{1/2}}{1-(1-4x)^{1/2}}\right], \tag{10.2.53}$$

with $K \to 0$ for $M_h \to 2m_e$ and $K \to 1$ for $M_h \gg m_e$. Therefore for this process the ratio of the lifetime to the Hubble time today $1/H_0$ is given by

$$H_0 \tau(\nu_h \to e^+ e^- \nu_m) \simeq \frac{10^{-13}}{|H_{eh}|^2 K\left[\frac{m_e^2}{M_h^2}\right]} \left(\frac{MeV}{M_h}\right)^5 . \tag{10.2.54}$$

The decay rate into active-like neutrinos mediated by neutral currents (not GIM (Glashow Iliopoulos Maiani)) suppressed with sterile-like heavy neutrinos) is given by (see [399])

$$\Gamma(\nu_h \to \nu_m \nu'_m \nu''_m) = 3.5 \times 10^{-5} \sum_{\alpha=e,\mu,\tau} |H_{\alpha h}|^2 \left(\frac{M_h}{MeV}\right)^5 \frac{1}{s} \tag{10.2.55}$$

therefore for this "invisible" channel we find

$$H_0 \tau(\nu_h \to \nu_m \nu'_m \nu''_m) \simeq \frac{10^{-13}}{\sum_{\alpha=e,\mu,\tau} |H_{\alpha h}|^2} \left(\frac{MeV}{M_h}\right)^5 . \tag{10.2.56}$$

The radiative decay $\nu_h \to \gamma \nu_m$ has been studied in ref.[414, 399, 28], for heavy (sterile-like) neutrinos it is not GIM suppressed but is suppressed by one power of α_{em} with respect to the leptonic channels above and given by

$$\Gamma(\nu_h \to \gamma \nu_m) \simeq 10^{-7} \Big(\frac{M_h}{\mathrm{MeV}}\Big)^5 \sum_\alpha |H_{h\alpha} U_{\alpha m}|^2 \frac{1}{s} \tag{10.2.57}$$

namely

$$H_0 \tau(\nu_h \to \gamma \nu_m) \simeq \frac{2 \times 10^{-11}}{\sum_\alpha |H_{h\alpha} U_{\alpha m}|^2} \left(\frac{MeV}{M_h}\right)^5 . \tag{10.2.58}$$

For a Majorana neutrino there is an extra factor 2 multiplying the rates for the expressions above. For the heavy neutrino to be stable during the lifetime of the Universe and be a

suitable DM candidate it must be that $H_0\tau(\nu_h \to e^+e^-\nu_m)$; $H_0\tau(\nu_h \to \nu_m\nu'_m\nu''_m)$; $H_0\tau(\nu_h \to \gamma\nu_m) \geq 1$.

Assuming that the $|H_{\alpha h}|$ are all of the same order and $|U_{\alpha m}| \simeq \mathcal{O}(1)$ these estimates suggest that heavy neutrinos with

$$|H_{h\alpha}|^2(M_h/MeV)^5 \lesssim 10^{-13} \tag{10.2.59}$$

feature lifetimes $\gtrsim 1/H_0$ and may be acceptable DM candidates. Interestingly, for $M_h \leq 1\,MeV$ the "visible" leptonic decay channel $\nu_h \to e^+e^-\nu_m$ shuts off and the lifetime is dominated by the "invisible" channel $\nu_h \to \nu_m\nu'_m\nu''_m$ therefore evading potential bounds from the "visible" decays into l^+l^- pairs.

As per the discussion above the constrain that the heavy neutrino features a lifetime $\tau > 1/H_0$ also implies that its distribution function is out of LTE. We will discuss possible interesting consequences of a heavy sterile neutrino with a lifetime much smaller than the Hubble time in a later section (see section10.5).

10.2.6 Comparisons and caveats

A formulation of the production rates of sterile neutrinos firmly based on the quantum field theory of neutrino mixing was introduced in ref.[390]. It relies on a see-saw type mass matrix with vanishing masses for the active neutrinos, large (Majorana) masses for sterile neutrinos on the diagonal and small off-diagonal matrix elements that mix the sterile and active neutrinos. This small (compared to the large Majorana masses) off-diagonal mixing sub-matrix is treated in a perturbative expansion. The authors in ref.[390] obtain the quantum density matrix by tracing over the degrees of freedom of the standard model (assumed in thermal equilibrium) up to second order in the off-diagonal mixing, and, in principle, to all orders in the strong and weak interactions. The quantum master equation that describes the time evolution of the reduced density matrix is completely determined by correlation functions of the active neutrino including self-energy corrections, which is written in terms of spectral representations. The production rate of sterile neutrinos is obtained from the imaginary part of this self energy evaluated on the mass-shell of the sterile neutrinos,

in principle to all orders in standard model couplings. The authors focus on temperature scales $\ll M_{W,Z}$, and argue that sterile neutrinos with masses in the keV range are primarily produced in the temperature range $T \sim 150\,\mathrm{MeV}$. Furthermore, they only consider a "gain" term in this temperature regime neglecting the loss term.

There are several differences between the approach of ref.[390] and the framework presented here: **1)** we describe the production process with standard quantum kinetic equations for the *mass eigenstates*, without resorting to any particular model for the mixing mass matrix, however, similarly to [390] our results are in principle valid to all orders in standard model couplings, but to second order in the (small) mixing matrix elements between the light (active-like) and heavy ("sterile-like") neutrinos. **2)** By going to the mass basis, we recognize that standard model processes that produce active neutrinos via charged and neutral current vertices lead to the production of heavy neutrinos provided the kinematics is favorable and identify various processes available in a wide temperature region. **3)** we consider both the gain and loss terms, giving an exact result for the non-equilibrium evolution of the distribution function, this is the result (10.2.36). We discuss the important issue of thermalization *vis a vis* the constraints of stability of the DM candidate.

The authors of ref.[390] identified the "one-loop" contributions to the active neutrino self energy, whose imaginary parts are precisely the contributions described by the quantum kinetic equations (10.2.13,10.2.14,10.2.17,10.2.18) but neglected them by restricting their study to $T \ll M_{W,Z}$, whereas we argue that these contributions may be important (obviously a statement on their impact requires a detailed calculation, beyond the scope of this article). The production processes mediated by W,Z exchange in an intermediate state with gain and loss terms given by (10.2.19-10.2.22) (and equivalent for charged current processes) correspond to the "two loops" contributions in ref.[390], although we pointed out that at $T \sim M_W, M_Z$ these processes could be resonantly enhanced as the intermediate W, Z propagators feature "on-shell" thermal contributions which are not suppressed at these temperatures. **4)** We have identified production processes from "collective excitations" in the medium, suggesting that finite temperature corrections, leading to plasmon masses for the photon and for the W,Z bosons *above* T_{EW} *may* yield important production mechanisms. **5)** we will argue below that different production channels with different kinematics yield different

contributions to the distribution function. This is a result of "kinematic entanglement" between the heavy neutrino and the lepton produced in the reaction leading to distribution functions that may be colder for some channels and warmer for others depending on the mass of the lepton and the kinematics. In other words, the final distribution is a *mixture* of several components even for the same species. This point will be elaborated upon in more detail within a specific example in the next sections.

Caveats: As we discussed above at high temperature there are important self-energy corrections that must be assessed for a more reliable understanding of the gain and loss processes. Furthermore, our results are valid up to $\mathcal{O}(|H_{\alpha h}|^2)$ under the assumption that $|H_{\alpha h}|^2 \ll |U_{\alpha m}|^2$, (and in principle to all orders in weak coupling), however this approximation will break down if there are Mikheyev-Smirnov-Wolfenstein (MSW)[30, 415, 416] resonances in the medium, because near the resonance the effective mixing angle reaches $\pi/4$. This is also a caveat in the approach of ref.[390] because in this reference the quantum master equation has been obtained up to second order in the sterile-active mixing angle.

Including the possibility of MSW resonances in the medium requires adopting a different framework that does not rely on $|H_{\alpha h}| \ll 1$ but that would be firmly based on quantum field theory out of equilibrium. Such approach *could be* based on effective field theory out of equilibrium as advocated in ref.[426]. We will report on this approach in a future study.

10.3 COSMOLOGICAL CONSEQUENCES AND CONSTRAINTS.

If there is a hierarchy of *stable* heavy neutrinos produced by the various mechanisms discussed above (we will comment later on unstable heavy neutrinos), each species with $M_h \gg 1\,\mathrm{eV}$ will contribute as a non-relativistic DM component after matter radiation equality. Once the distribution function at freeze out is obtained, various cosmological consequences of the dark matter species can be assessed. In this section we gather together various quantities of cosmological relevance which are determined by basic properties of the dark matter species: mass, number of intrinsic degrees of freedom, distribution function and freeze-out (or de-

coupling) temperature with the purpose of assessing particular species of heavy neutrinos as DM candidates once their distribution function is obtained.

Let us define the *total* asymptotic "frozen" distribution function of a species ν_h of mass M_h as

$$f_h(q_c) = n_h(q_c; \infty)\,, \tag{10.3.1}$$

where by $t \to \infty$ we mean a time sufficiently long that the distribution function satisfies the freeze-out condition (10.2.39).

As discussed above after freeze-out the distribution function depends on the physical momentum $q_f(t) = q_c/a(t)$ and temperature $T(t)$ through the combination $y = q_f(t)/T(t) = q_c/T_0$ where T_0 is the temperature at which the species decoupled from the plasma (freeze-out) redshifted to today. Through the usual argument of entropy conservation it is related to the temperature of the cosmic microwave background today $T_{\gamma,0}$ by

$$\frac{T_0}{T_{\gamma,0}} = \Big(\frac{2}{g_{dh}}\Big)^{1/3}\,, \tag{10.3.2}$$

where g_{dh} is the number of ultrarelativistic degrees of freedom at the time when the particular species ν_h decoupled (freezes). Then the number density $\mathcal{N}_{\nu_h}$, energy density ρ_{ν_h} and *partial pressure* $\mathcal{P}_{\nu_h}$ of species ν_h is given by

$$\mathcal{N}_{\nu_h}(t) = g_{\nu_h}\int \frac{d^3q_f}{(2\pi)^3} f_h(q_c) = \frac{g_{\nu_h}}{2\pi^2}\Big(\frac{T_0}{a(t)}\Big)^3 \int_0^\infty dy\, y^2 f_h(y) \tag{10.3.3}$$

$$\rho_{\nu_h}(t) = g_{\nu_h}\int_0^\infty \frac{d^3q_f}{(2\pi)^3}\sqrt{q_f^2+M_h^2}\, f_h(q_c) = \frac{g_{\nu_h}M_h}{2\pi^2}\Big(\frac{T_0}{a(t)}\Big)^3 \int_0^\infty dy\, y^2\sqrt{1+\frac{y^2}{x_h^2(t)}}\, f_h(y) \tag{10.3.4}$$

$$\mathcal{P}_{\nu_h}(t) = \frac{g_{\nu_h}}{3}\int_0^\infty \frac{d^3q_f}{(2\pi)^3}\frac{|\vec{q}_f|^2}{\sqrt{q_f^2+M_h^2}} f_h(q_c) = \frac{g_{\nu_h}}{6\pi^2 M_h}\Big(\frac{T_0}{a(t)}\Big)^5 \int_0^\infty dy\, \frac{y^4\, f_h(y)}{\sqrt{1+\frac{y^2}{x_h^2(t)}}} \tag{10.3.5}$$

where g_{ν_h} is the number of internal degrees of freedom of the species ν_h, and

$$x_h(t) = M_h/T(t)\,. \tag{10.3.6}$$

The contribution of each non-relativistic species ν_h to Ω_{DM} today with $T_0/M_h \ll 1$ is given by

$$\Omega_{\nu_h} h^2 = \frac{g_{\nu_h} M_h}{2\pi^2 \rho_c} T_0^3 \, h^2 \int_0^\infty dy \, y^2 f_h(y) = \Big(\frac{n_\gamma \, h^2}{\rho_c}\Big) \frac{g_{\nu_h} M_h}{4 \, \xi(3)} \Big(\frac{T_0}{T_{\gamma,0}}\Big)^3 \int_0^\infty dy \, y^2 f_h(y) \, . \quad (10.3.7)$$

Using the relation (10.3.2) and $n_\gamma h^2/\rho_c = 1/25.67 \, eV$ we find

$$\Omega_{\nu_h} h^2 = \frac{M_h}{61.7 \, eV} \Big(\frac{g_{\nu_h}}{g_{dh}}\Big) \int_0^\infty dy \, y^2 f_h(y) \, . \quad (10.3.8)$$

Therefore, assuming that all of DM is in the form of various species of heavy neutrinos, it follows that

$$\Omega_{DM} h^2 = \sum_{h=4,5\cdots} \Big(\frac{M_h}{61.7 \, eV}\Big) \Big(\frac{g_{\nu_h}}{g_{dh}}\Big) \int_0^\infty dy \, y^2 f_h(y) \, . \quad (10.3.9)$$

The *fraction* $\mathcal{F}_{\nu_h}$ of Ω_{DM} contributed by the species ν_h is given by

$$\mathcal{F}_{\nu_h} \equiv \frac{\Omega_{\nu_h} h^2}{\Omega_{DM} h^2} = \frac{M_h}{7.4 \, eV} \Big(\frac{g_{\nu_h}}{g_{dh}}\Big) \int_0^\infty dy \, y^2 f_h(y) \, , \quad (10.3.10)$$

where we used $\Omega_{DM} h^2 = 0.1199$[334].

The equation of state for each species is

$$w_{\nu_h}(T(t))) = \frac{\mathcal{P}_{\nu_h}(t)}{\rho_{\nu_h}(t)} = \frac{1}{3} \frac{\int_0^\infty dy \, \frac{y^4 \, f_h(y)}{\sqrt{y^2 + \frac{M_h^2}{T^2(t)}}}}{\int_0^\infty dy \, y^2 \sqrt{y^2 + \frac{M_h^2}{T^2(t)}} \, f_h(y)} \, . \quad (10.3.11)$$

The equation of state yields information on how "cold" is the particular species and provides a benchmark to compare the equilibrium and non-equilibrium distribution functions. In particular $\sqrt{w_{\nu_h}(T)}$ yields a generalization of the effective adiabatic "speed of sound" for collisionless DM, namely, $\mathcal{P}_{\nu_h}(t) = c_h^2(T(t)) \, \rho_{\nu_h}(t) \;\; ; \;\; c_h^2(T(t)) \equiv w_{\nu_h}(T(t))$.

In the non-relativistic limit $M_h \gg T(t) \;\; ; \;\; x(t) \to \infty$, the average primordial velocity dispersion for a species ν_h is given by

$$\big\langle \vec{V}_{\nu_h}^2(t) \big\rangle = \frac{\langle \vec{q}_f^{\,2} \rangle}{M_h^2} = \frac{T(t)^2}{M_h^2} \frac{\int dy \, y^4 f_h(y)}{\int dy \, y^2 f_h(y)} = \frac{3 \mathcal{P}_{\nu_h}}{\rho_{\nu_h}} \quad (10.3.12)$$

which leads to the *primordial* velocity dispersion relation

$$\mathcal{P}_{\nu_h} = \sigma^2_{\nu_h}\,\rho_{\nu_h} \quad ; \quad \sigma_{\nu_h} = \sqrt{\frac{\langle \vec{V}^2_{\nu_h}(t)\rangle}{3}} = \frac{T(t)}{M_h}\sqrt{\frac{\int dy\, y^4 f_h(y)}{3\int dy\, y^2 f_h(y)}}\,, \tag{10.3.13}$$

therefore for a collisionless non-relativistic species ν_h the dispersion $\sigma_{\nu_h} = w^{1/2}_{\nu_h}$ plays the role of the adiabatic sound speed. The equation of state (10.3.11) will be important in assessing the consequences of kinematic entanglement of distribution functions obtained by different production channels since it yields information on the "coldness" of the distribution.

In analogy with a fluid description and following ref.[427] we introduce the *comoving* free streaming wave vector for the species ν_h similarly to the comoving Jeans wavevector, namely

$$k^2_{fs,\nu_h}(t) = \frac{4\pi\, G\,\rho_m(t)}{\langle \vec{V}^2_{\nu_h}(t)\rangle}\, a^2(t) \tag{10.3.14}$$

as discussed in ref.[427] $k_{fs}(t_{eq})$ describes the cutoff in the power spectrum for linear density perturbations, where t_{eq} is the time of matter-radiation equality. Since for non-relativistic particles

$$\langle \vec{V}^2_{\nu_h}(t)\rangle = \frac{\langle \vec{V}^2_{\nu_h}(0)\rangle}{a^2(t)} \tag{10.3.15}$$

where $\langle \vec{V}^2_{\nu_h}(0)\rangle$ is the velocity dispersion today, it follows that

$$k_{fs,\nu_h}(t_{eq}) = k_{fs,\nu_h}(0)\,\sqrt{a(t_{eq})} \tag{10.3.16}$$

where

$$k_{fs,\nu_h}(0) = \left[\frac{3\Omega_M h^2}{2\vec{V}^2_{\nu_h}(0)}\right]^{1/2} H_0 \tag{10.3.17}$$

is the free streaming wavevector *today*. The free streaming scale that determines the length scale associated with the cutoff in the power spectrum is

$$\lambda_{fs,\nu_h}(t_{eq}) = \frac{2\pi}{k_{fs,\nu_h}(t_{eq})} = \lambda_{fs,\nu_h}(0)\,(1+z_{eq})^{1/2} \tag{10.3.18}$$

with $\Omega_M h^2 = \Omega_b h^2 + \Omega_{DM} h^2 = 0.14$ we find[2]

$$\lambda_{fs,\nu_h}(0) = 9.72\,\left(\frac{keV}{M_h}\right)\left(\frac{2}{g_{dh}}\right)^{1/3}\sqrt{\frac{\int dy\, y^4 f_h(y)}{\int dy\, y^2 f_h(y)}}\ \text{kpc} \tag{10.3.19}$$

[2]The discrepancy with the result in ref.[396] is due to the difference between $\Omega_{DM}h^2 = 0.12$ and $\Omega_M h^2 = 0.14$[334].

and $(1+z_{eq})^{1/2} \simeq 57$.

Up to constants of $\mathcal{O}(1)$) $\lambda_{fs,\nu_h}(t_{eq})$ is equivalent to the comoving distance traveled by a free streaming particle with average velocity $\left\langle \vec{V}^2_{\nu_h}(t) \right\rangle$ from the time of matter-radiation equality until today[427], as can be seen from the following argument. Consider a particle moving with velocity $v(t) = \sqrt{\left\langle \vec{V}^2_{\nu_h}(t) \right\rangle}$ the comoving distance traveled between t_{eq} and today at time $t^* \gg t_{eq}$ is

$$l(t_{eq}, t^*) = \int_{t_{eq}}^{t^*} \frac{v(t')}{a(t')} dt' \tag{10.3.20}$$

during a matter dominated cosmology when density perturbations grow

$$a(t) = \left[\Omega_M H_0^2\right]^{1/3} t^{2/3} \tag{10.3.21}$$

and using (10.3.15) for a non-relativistic species we find

$$l(t_{eq}, t^*) \simeq 3 \left[\frac{\left\langle \vec{V}^2_{\nu_h}(0) \right\rangle}{\Omega_M H_0^2} \right]^{1/2} (1+z_{eq})^{1/2} \,, \tag{10.3.22}$$

from which it follows that

$$\lambda_{fs,\nu_h}(t_{eq}) \simeq 1.7\, l(t_{eq}, t^*) \,. \tag{10.3.23}$$

Phase-space density (Tremaine-Gunn) constraints are based on the result that the coarse grained phase space density may only decrease from its primordial value during "violent relaxation" in the process of galaxy formation and evolution[382, 383, 385, 322, 359]. The coarse grained phase space density for the species ν_h is defined as[322]

$$\mathcal{D}_h = \frac{\mathcal{N}_h(t)}{\left[\langle q_f^2(t) \rangle_h\right]^{3/2}} \,, \tag{10.3.24}$$

where the average is with the distribution function $f_h(q_c)$ and $\mathcal{N}_h(t)$ is given by (10.3.3). Therefore the primordial coarse grained phase space density is given by

$$\mathcal{D}_h = \frac{g_{\nu_h}}{2\pi^2} \frac{\left[\int_0^\infty dy\, y^2 f_h(y)\right]^{5/2}}{\left[\int_0^\infty dy\, y^4 f_h(y)\right]^{3/2}} \,. \tag{10.3.25}$$

When the DM particles become non relativistic it becomes

$$\mathcal{D}_h = \frac{\rho_{\nu_h}}{M_h^4 \left\langle \vec{V}^2_{\nu_h} \right\rangle^{3/2}} = \frac{1}{3^{3/2} M_h^4} \frac{\rho_{\nu_h}}{\sigma^3_{\nu_h}} \tag{10.3.26}$$

For a primordial phase space density, $\mathcal{D}_h$, imposing the bound $\mathcal{D}_{today} \leq \mathcal{D}_h$[322] gives us the constraint

$$\mathcal{D}_h \geq \frac{1}{3^{3/2} M_h^4} \left[\frac{\rho_{DM}}{\sigma_{DM}^3}\right]_{today} \tag{10.3.27}$$

If there is only *one species* the values of ρ_{ν_h} and σ_{ν_h} can be inferred from the kinematics of dwarf spheroidal galaxies[322] and this constraint can be used to obtain a bound which complements those from CMB observations.

10.3.1 Non-LTE freeze-out from multiple production channels:

The distribution function $f_h(y)$ is a result of *all* the possible production channels that are kinematically allowed as the total gain and loss terms determine the solution of the cosmological quantum kinetic equation. Furthermore, each production process of a heavy neutrino species ν_h may actually be the result of the decay of a heavy standard model particle (such as $W, Z, \tau, \mu, \cdots$) into different channels and each channel may yield a different distribution function because of the kinematics, we refer to this as "kinematic entanglement". Thus even for a single species ν_h of heavy neutrino, its frozen distribution function may be a *mixture* of several contributions some colder than others as a consequence of the kinematics.

While the general solution (10.2.36) is *non-linear* in the gain and loss terms of each channel because of the exponential terms in the relaxation rate γ, *if* the condition (10.2.40) is fulfilled, these exponentials can be neglected and the result is given by (10.2.41) which is linear in the gain rates allowing for an identification of the distribution functions associated with *each production channel*. In each production channel the heavy neutrino is *kinematically entangled* with a lepton and the non-LTE distribution function at freeze out will reveal this "entanglement" for example in the form of production thresholds. This aspect is another manifestation of "mixed DM", in the sense that even for a given heavy species ν_h, its distribution function at freeze-out is a result of different contributions from different channels each with different kinematics. The concentration of each component depends, among other factors, of the ratios of mixing angles for the different channels. If freeze-out occurs out of

LTE, the distribution functions will maintain memory of the processes that led to the production and the kinematic entanglement, in LTE this memory is erased as the distribution function becomes the Fermi-Dirac distribution regardless of the production process.

Separating the contribution from the different channels will also allow an assessment of the "coldness" of the heavy neutrino as a result of the particular production channel. An explicit example of these phenomena will be studied in detail within the context of heavy neutrinos produced from pion decay in the next section.

10.3.2 Summary of cosmological constraints:

We now summarize the main cosmological constraints in terms of the decoupled distribution function, degrees of freedom and mass of a particular heavy species ν_h. Once the distribution function has been obtained from the solution of the quantum kinetic equations these constraints inform the feasibility of such species as a suitable DM candidate. In the discussion below $f_h(y)$ is the *total distribution function* solution of (10.2.36) after freeze-out.

- **Abundance:** the fraction of DM from a particular species ν_h must obey, $\mathcal{F}_{\nu_h} \leq 1$ leading to the abundance constraint

$$\frac{M_h}{7.4\,eV}\left(\frac{g_{\nu_h}}{g_{dh}}\right)\int_0^\infty dy\, y^2 f_h(y) \leq 1\,, \tag{10.3.28}$$

- **Phase space density (Tremaine-Gunn):** the result that the phase space density only diminishes during "violent relaxation" in gravitational collapse[382, 383, 385, 322] leading to (10.3.27) yields the constraint

$$0.26\, g_{\nu_h}\left[\frac{M_h}{keV}\right]^4 \frac{\left[\int_0^\infty dy\, y^2 f_h(y)\right]^{5/2}}{\left[\int_0^\infty dy\, y^4 f_h(y)\right]^{3/2}} > \left[\frac{\rho_{DM}}{\sigma^3_{DM}\,(keV)^4}\right]_{today}\,, \tag{10.3.29}$$

or alternatively[3]

$$0.335\times 10^8\, g_{\nu_h}\left[\frac{M_h}{keV}\right]^4 \frac{\left[\int_0^\infty dy\, y^2 f_h(y)\right]^{5/2}}{\left[\int_0^\infty dy\, y^4 f_h(y)\right]^{3/2}} > \left[\frac{\rho_{DM}}{M_\odot/kpc^3}\right]_{today} \times \left[\frac{(km/s)^3}{\sigma^3_{DM}}\right]_{today}\,. \tag{10.3.30}$$

[3] $1\,keV^4 = 1.27\times 10^8 \frac{M_\odot}{kpc^3}\,(\frac{km}{s})^{-3}$.

The most DM dominated dwarf spheroidal galaxies feature phase space densities within a wide range (see [359] and references therein)

$$10^{-4}\,(keV)^4 \lesssim \frac{\rho}{\sigma^3} \lesssim 1\,(keV)^4\,. \tag{10.3.31}$$

- **Stability:** To be a DM candidate the candidate particle must feature a lifetime larger than the Hubble time $1/H_0$, namely $\tau H_0 > 1$. Heavy neutrinos feature various leptonic and radiative decay channels analyzed in section (10.2). A conservative bound on the lifetime from the dominant leptonic decay is (see eqns. (10.2.52-10.2.59))

$$|H_{h\alpha}|^2 (M_h/MeV)^5 \lesssim 10^{-13}\,, \tag{10.3.32}$$

in particular for heavy neutrinos with $M_h < 1\,\text{MeV}$ the decay channel with the largest branching ratio corresponds to the "invisible" decay into three light active-like neutrinos. This decay mode provides interesting possibilities *even when the heavier neutrino is unstable*, this will be discussed in section (10.5) below.

Caveats: If indeed DM is composed of several species including a hierarchy of different masses, the individual free streaming lengths and phase space densities of a species ν_h of mass M_h may *not* be the proper characterization. Although the different species do not interact directly with each other (or do so extremely feebly), they interact indirectly via the common gravitational potential which is sourced by all species. While a previous study addressed the gravitational clustering properties of mixed dark matter[386] that study did not include cosmological expansion and should be deemed, at best, as of preliminary guidance. As far as we know, there has not yet been a consistent study of free streaming and phase space dynamics for mixed DM including cosmological expansion during the different stages. In particular on the important question of what is the correct cut-off scale in the linear power spectrum in the case of various components. This entails the study of the linearized collisionless Boltzmann equation including cosmological expansion. While a numerical study implementing the public Boltzmann codes may yield insight, to the best of our knowledge an analytical study for arbitrary concentrations of the different components clarifying the

contributions of each species to the effective cutoff scale is still lacking. Until such study emerges, we will consider the free-streaming and phase space density constraints of the individual species as *indicative*.

10.4 HEAVY NEUTRINO PRODUCTION FROM PION DECAY.

The previous section discusses in general terms the quantum kinetic approach to production and freeze-out of heavy neutrino species and the various processes that may produce them in the early Universe. It is clear from the discussion in section (10.2) that standard model interaction vertices that lead to the production of active neutrinos will also lead to the production of the heavy neutrinos as long as the processes are kinematically allowed. This entails a far broader range of production mechanisms than those that had been the focus in the literature and suggests that a firm assessment of a particular candidate, such as a keV "sterile" neutrino requires a deeper understanding of *all* the standard model processes that may lead to their production during the various cosmological stages.

We recognized the necessity to include finite temperature corrections to masses and interaction vertices to obtain a reliable description. In this section we implement this program with a definite example: the production of heavy neutrinos from pion decay shortly after the QCD phase transition (crossover) into the confined phase.

In ref.[390] the authors proposed to study hadronic contributions to the production of sterile neutrinos by considering the self-energy corrections to the active (massless) neutrinos in terms of correlators of vector and axial-vector currents. While this is correct in principle, it is an impractical program: the confined phase of QCD is strongly coupled and the description in terms of nearly free quarks is at best an uncontrolled simplification, casting doubts on the reliability of the conclusions in this reference.

Instead, in this section we study the production from pion decay by relying on the well understood effective field theory description of weak interactions of pions, enhanced by the results of a systematic program that studied finite temperature corrections to the pion

decay constant and mass implementing non-perturbative techniques from chiral perturbation theory, linear and non-linear sigma models and lattice gauge theory [370, 371, 373, 374, 372, 375, 376, 377, 369]. There are at least three important reasons that motivate this study: **1)** it is a clear and relevant example of the quantum kinetic equation approach to production and freeze out that includes consistently finite temperature corrections. In fact this case is similar to the production via the decay of W, Z vector bosons the main difference being the three polarizations of the latter and kinematic factors. **2)** Pion decay surely contributed to the production of heavy neutrinos in the early Universe if such species of neutrinos do exist: the QCD transition to its confined phase undoubtedly happened, and the consequent hadronization resulted in baryons and low lying mesons, pions being the lightest, are the most populated meson degrees of freedom near the QCD scale. **3)** The two leptonic decay channels $\pi \to \mu\nu_h; \pi \to e\nu_h$ feature different kinematics and thresholds, in particular the (V-A) vector-axial coupling results in chiral suppression of the e channel for vanishingly small M_h. Since the heavy neutrino produced in the decay is *kinematically entangled* with the emitted lepton[139], we expect that the distribution functions from each channel will display differences as a manifestation of this kinematic entanglement, thus providing an explicit example of memory effects as a consequence of "kinematic entanglement" in the non-equilibrium distribution function and dependence on the particular production channel.

At temperatures larger than the QCD Phase transition $T_{QCD} \sim 155 MeV$[369], quarks and gluons are asymptotically free. Below this temperature, QCD bound states form on strong interaction time scales, the lightest being the pion. Recent lattice gauge theory calculations [369] suggest that the confinement-deconfinement transition is not a sharp phase transition but a smooth yet rapid crossover at a temperature $T_c \sim 155 MeV$ within a temperature range $\Delta T \pm 10 MeV$. This occurs in the radiation dominated epoch at $t \simeq 10\,\mu secs$ within a time range $\Delta t \sim 2 - 3\,\mu secs$. This is *much* larger than the typical strong and electromagnetic interaction time scales $\simeq 10^{-22}\,secs$ implying that pions that form shortly after the confining cross-over are brought to LTE via strong, electromagnetic and weak interactions on time scales much shorter than Δt. After pions are produced they reach LTE via $\pi - \pi$ scattering on strong interaction time scales, they decay into leptons and maintain LTE via detailed balance with the inverse process since charged leptons and active-like

neutrinos are also in LTE. In conclusion, for $T \lesssim 155 MeV$, pions are present in the plasma in thermal/chemical equilibrium due to pions interacting on strong interaction time scales ($10^{-22} s$) while the crossover transition occurs on the order of $2 - 3\,\mu secs$.

If the pions (slowly) decay into heavy neutrinos ν_h with very small mixing angles, detailed balance for $\pi \rightleftharpoons l\nu_h, h = 4, 5 \cdots$ will *not* be maintained if the heavy neutrinos are *not* in LTE. As the pion is the lowest lying bound state of QCD, it is a reasonable assumption that during the QCD confinement-deconfinemet crossover pions will be the most dominantly produced mesonic bound state and, during this time, pions will remain in LTE with the light *active-like* neutrinos and charged leptons by detailed balance $\pi \rightleftharpoons l_\alpha \nu_\alpha$. We focus on heavy neutrino ν_h production from $\pi \to l\nu_h$ which is *suppressed* by $|H_{\alpha h}|^2 \ll 1$ with respect to the active neutrinos and does not maintain detailed balance.

A full study of sterile neutrino production through π decay in the early universe requires various finite temperature corrections and a preliminary study which focused on the production of neutrinos in the keV mass range[396] has implemented the first step. This study yielded a suitable warm dark matter candidate with free streaming lengths on the order of several kpc whereas it is expected that heavier neutrinos will yield a colder dark matter species.

The issue of a heavy neutrino production ($M_h \gtrsim MeV$) through π decay has not been addressed and is the main focus of this section. Through the two possible channels $\pi \to \mu\nu_\mu; \pi \to e\nu_e$ pion decay offers a wide kinematic window to the production of heavy neutrinos and provides a natural mechanism to produce a *mixed dark matter* scenario provided that a hierarchy of heavy neutrino species exist and their production is kinematically allowed[428, 386]. Furthermore, "kinematic entanglement" will yield different distribution functions for the different channels, thus providing an example of a *mixed* distribution for the same species.

Whereas pion decay $\pi \to \mu\nu$, is one of the most ubiquitous mechanisms to produce neutrinos in many terrestrial experiments, including short baseline experiments, this production mechanism has not been fully addressed within the cosmological setting. While we *do not* claim that the processes studied below are more or less important than the processes described in the previous section or discussed elsewhere in the literature, this study leads to

a clear example of the concepts and methods described in the previous section, including the finite temperature corrections. Furthermore, this production mechanism *may* yield a substantial (if not the dominant) contribution to DM below the QCD scale.

In studying the production of heavy neutrinos, we make the following assumptions for the quantum kinetic equation that governs the ν_h population build up:

- A substantial body of work has investigated finite temperature corrections to the pion decay constant [373, 372, 377, 374, 376]. To leading order in chiral perturbation theory, the correction is given by

$$f_\pi^2 \to f_\pi^2(t) = f_\pi^2(0)\left(1 - \frac{T(t)^2}{6 f_\pi(0)^2}\right) \quad ; \quad f_\pi(0) = 93 MeV\,. \tag{10.4.1}$$

  This result will be needed for the quantum kinetics equation as we consider production beginning at $T_{QCD} \sim 155 MeV$. These results are well established but this subfield remains an active area of investigation [429] and, for precision studies, these results will need refinement. The assumption is that there are no pions present in the plasma above $T_{QCD} \sim 155 MeV$ and that hadronization happens instantaneously in comparison to the neutrino production time scales.
- Finite temperature corrections to the masses of both the pions and charged leptons occur in the plasma. The pion mass including electromagnetic corrections is shown in figure 2 of [372]. It is shown that between the temperatures of 10 and 150 MeV, the pion mass varies between 140 and 144 MeV. This change is relatively small and will be neglected for the calculation and we will use an average $M_\pi = 142 MeV$. Finite temperature corrections to the charged lepton mass are of $\mathcal{O}(eT)$[418, 419, 421, 417] and for muons this correction is only a fraction of its mass so it can be neglected. However, mass corrections to the electron may be substantial, but in this case we are interested in a kinematic window of large mass for ν_h. For the purposes of this work, the charged lepton mass will be taken as a constant for both muons and electrons. The effects of charged lepton mass corrections will be investigated elsewhere.
- It is argued in [430] that the chemical potentials (including pions) are on the order of $\sim 10^{-6} eV$ for the temperature range we consider here. We assume that the lepton

asymmetry is small and will be neglected in our calculations consistently with the neglect of chemical potentials in the quantum kinetic equations of the previous section.

- The mixing angle between active-sterile neutrinos in the presence of a matter potential will develop a temperature and lepton asymmetry dependence as shown in refs. [415, 40, 346, 416]. For vanishing lepton asymmetries (chemical potentials) and $T/M_W \ll 1$ there are no in medium MSW resonances[415, 40, 416]. In absence of lepton asymmetry the finite temperature in-medium potential lead to the following modification of the mixing angles (we consider mixing between one active-like and a heavy neutrino only)

$$\sin(2\theta_m) = \frac{\sin 2\theta}{\left[\sin^2 2\theta + (\cos 2\theta + V^{th}/\Delta)^2\right]^{1/2}} \tag{10.4.2}$$

where $\Delta = \delta M^2/2E = (M_h^2 - M_a^2)/2E$, and $V^{th} \sim 10^2 G_F^2 E T^4$[415, 40]. We will be concerned with the temperature range $E \sim T \sim 100\, MeV$, we find

$$V^{th}/\Delta \sim 10^{-2} \left(\frac{T}{100\,\mathrm{MeV}}\right)^6 \left(\frac{\mathrm{keV}}{M_h}\right)^2 \tag{10.4.3}$$

The focus of this study is mainly on neutrinos with $M_h \gtrsim$ few keV and $\sin(2\theta) \ll 1$ at temperatures below $T \sim 150 MeV$ so the temperature dependence of the mixing can be ignored. In any event, the temperature correction to the mixing angles may, at most, lead to a slight quantitative change but not to a major modification of our main results, in particular the contributions from different channels with kinematic entanglement is a robust feature, as it will become clear from the analysis below.

The effective low energy field theory that describes pion decay appended by finite temperature corrections to the $\pi^{\pm}$ mass and decay constant yields the following interaction Hamiltonian (in the interaction picture),

$$H_I = \sum_{\alpha=e,\mu} \sqrt{2}\, G_F\, V_{ud}\, f_\pi(T) \int d^3x\, [\bar{\nu}_\alpha(\vec{x},t)\gamma^\sigma \mathcal{P}_L \Psi_\alpha(\vec{x},t) J_\sigma^\pi(\vec{x},t) + h.c.] \tag{10.4.4}$$

where $J_\sigma^\pi = i\partial_\sigma \pi(x,t)$ is the pseudoscalar pion current, G_F is the Fermi constant, V_{ud} is the CKM matrix element, f_π(T) is the pion decay constant with finite temperature corrections

given by (10.4.1) and the "flavor" neutrino fields ν_α are related to the fields that create-annihilate neutrino mass eigenstates by the relation (10.2.1).

Pions decay into heavy neutrinos through the channels $\pi^+ \to \mu^+ \nu_h$; $\pi^+ \to e^+ \nu_h$ (and the charge conjugate). Because of the kinematics the heavy neutrino ν_h is "entangled" with the charged lepton l in the sense that the gain and loss rates depend on the particular channel. Therefore for each channel we label the gain and loss rates and the distribution function with the labels l, h. As discussed in detail in section (10.2) the total rates are the sum over all channels and the quantum kinetic equation inputs the total rates as per eqn. (10.2.34)

The quantum kinetic equation for ν_h production is given by (10.2.33) with the total gain and loss rates summed over the kinematically allowed channels labeled by the corresponding charged lepton l. The results for these rates are given in detail in appendix (10.7.1),

$$\begin{aligned} \Gamma^<_{lh}(q) &= \frac{|H_{lh}|^2 |V_{ud}|^2 G_F^2 f_\pi^2(T)}{8\pi} \frac{M_\pi^2 (M_l^2 + M_h^2) - (M_l^2 - M_h^2)^2}{q E_h(q)} \\ &\times \int_{p_-}^{p_+} \frac{dp\, p}{\sqrt{p^2 + M_\pi^2}} \Big[n_\pi(p)(1 - n_{\bar{l}}(|\vec{p} - \vec{q}|)) \Big] \end{aligned} \tag{10.4.5}$$

and

$$\Gamma^>_{lh}(q) = e^{E_h(t)/T(t)}\, \Gamma^<_{lh}(q)\,, \tag{10.4.6}$$

where momenta and energies are replaced by their local expressions in the expanding cosmology (10.2.30) and

$$p_\pm(t) = \left| \frac{E_h(q,t)}{2M_h^2} [(M_\pi^2 - M_l^2 + M_h^2)^2 - 4M_\pi^2 M_h^2]^{1/2} \pm \frac{q_f(t)(M_\pi^2 - M_l^2 + M_h^2)}{2M_h^2} \right| . \tag{10.4.7}$$

and the constraint from the energy conserving delta function

$$E_\pi(p,t) = E_l(k,t) + E_h(q,t) \quad ; \quad k = |\vec{p} - \vec{q}|\,. \tag{10.4.8}$$

As argued previously, the pions and charged leptons $\pi^\pm, l^\pm$ are assumed to be in LTE so that their distributions will take the standard bosonic and fermionic forms

$$n_{\pi^\pm} = \frac{1}{e^{E_\pi(p,t)/T} - 1} \quad ; \quad n_{\bar{l}/l} = \frac{1}{e^{E_l(k,t)/T} + 1} \quad ; \quad E(q,t) = \sqrt{\frac{q_c^2}{a(t)^2} + M^2} \tag{10.4.9}$$

where q_c is the comoving momentum and we neglect chemical potentials in this study.

Inserting the distributions into eqn. (10.4.5) and using $E_l(k) = E_\pi(p) - E_h(q)$, the gain rate becomes

$$\begin{aligned}\Gamma^{<}_{lh}(q) &= \frac{|H_{ls}|^2|V_{ud}|^2 f_\pi(T(t))^2}{16\pi}\frac{\left[M_\pi^2(M_l^2+M_h^2)-(M_l^2-M_h^2)^2\right]}{q\sqrt{q^2+M_h^2}} \\ &* \int_{p_-(t)}^{p_+(t)} \frac{dp\,p}{\sqrt{p^2+M_\pi^2}}\left[\frac{e^{-E_\nu(q)/T}e^{E_\pi(p)/T}}{(e^{E_\pi(p)/T}-1)(e^{-E_\nu(q)/T}e^{E_\pi(p)/T}+1)}\right],\end{aligned} \tag{10.4.10}$$

where $p^\pm$ are the solutions to the phase space constraints of eqn. (10.7.17) and the time dependence of the physical momentum has been suppressed. This integral is readily solved by a substitution with the following result

$$\begin{aligned}\Gamma^{<}_{lh}(q) &= \frac{|H_{lh}|^2|V_{ud}|^2 f_\pi^2(t)}{16\pi}\frac{\left[M_\pi^2(M_l^2+M_h^2)-(M_l^2-M_h^2)^2\right]}{q(t)E_h(q,t)(e^{E_h(q,t)/T(t)}+1)}T(t) \\ &* \ln\left(\frac{e^{E_\pi/T(t)}-1}{e^{-E_h(q,t)/T(t)}e^{E_\pi/T(t)}+1}\right)\Bigg|_{E_\pi=E_\pi^-(q_f(t))}^{E_\pi=E_\pi^+(q_f(t))}.\end{aligned} \tag{10.4.11}$$

where

$$E_\pi^\pm(q_f(t)) = \frac{1}{2M_h^2}\left[E_h(q_f(t))\Big(M_\pi^2+M_h^2-M_l^2\Big) \pm q_f(t)\lambda(M_\pi,M_l,M_h)\right] \tag{10.4.12}$$

and the threshold function is defined as

$$\lambda(M_\pi,M_l,M_h) = \left[M_\pi^4+M_l^4+M_h^2-2M_\pi^2M_l^2-2M_\pi^2M_h^2-2M_h^2M_l^2\right]^{1/2}. \tag{10.4.13}$$

In the limit $M_h \to 0$ the bracket in the first line of (10.4.11) yields the usual factor $M_l^2(M_\pi^2 - M_l^2)$, which is the hallmark of pion decay vanishing in the $M_l \to 0$ limit.

In eqn. (10.4.11), it is readily seen that $\Gamma^<$ depends on the ratio $y = \frac{q_f(t)}{T(t)} = \frac{q_c}{T_0}$ where T_0 the temperature of the plasma *today* with the scale factor normalized today at $(t = t_*)$ with $a(t_*) = 1$ and q_c is the comoving momentum.

The threshold function is one of the signatures of "kinematic entanglement"; for fixed M_l, the threshold function vanishes at $M_\nu = M_\pi - M_l$, for this value of M_h the two roots $E_\pi^\pm$ coalesce and the rate vanishes. This is important because for $M_h \lesssim 36\,\mathrm{MeV}$ there are *two* production channels of heavy neutrinos, $\pi \to \mu\nu_h; \pi \to e\nu_h$, whereas for $M_h > 36\,\mathrm{MeV}$ only $\pi \to e\nu_h$ is kinematically available.

Restricting our study to the production process $\pi \to l\nu_h$ and its inverse, we anticipate that the condition for production and freeze-out out of LTE (10.2.40) will be fulfilled, this will be proven self-consistently below. In this case the solution of the quantum kinetic equation is given by (10.2.41), namely the quantum kinetic equation (10.2.33) simplifies to

$$\frac{dn_h(q_c;t)}{dt} = \Gamma^<_{tot}(q_c;t)\,, \tag{10.4.14}$$

where

$$\Gamma^<_{tot}(q_c;t) = \sum_{l=\mu,e} \Gamma^<_{lh}(q_c;t) \tag{10.4.15}$$

and the sum is over the kinematically allowed channels. In this approximation the linearity of (10.4.14) allows us to introduce a distribution function for *each channel* $n_{lh}(q_c;t)$ that obeys

$$\frac{dn_{lh}(q_c;t)}{dt} = \Gamma^<_{lh}(q_c;t)\,, \tag{10.4.16}$$

which will prove useful as it allows the opportunity to understand how the "kinematic entanglement" associated with each channel is manifest in the frozen distribution function.

In this approximation the total distribution function is

$$n_h(q_c;t) = \sum_{l=\mu,e} n_{lh}(q_c;t)\,. \tag{10.4.17}$$

and after freeze-out

$$f_h(q_c) = n_h(q_c;t=\infty) = \sum_{l=\mu,e} f_{lh}(q_c;t=\infty)\,, \tag{10.4.18}$$

in other words the total distribution function after freeze-out is a *mixture* of the contributions from the different channels.

We emphasize that whereas the above definition is useful to learn the effects of the "kinematic entanglement" in the frozen distribution function, the cosmologically relevant quantities, such as coarse-grained phase space densities, free streaming length etc, are *non-linear* functions of the distribution function as they involve moments, therefore for these quantities only the total distribution function n_h (10.4.17) is relevant.

It proves convenient to introduce a dimensionless parameter τ that will play the role of time along with with a change of variables that introduces manifestly a factor of the Hubble factor (during radiation domination):

$$y = \frac{q_f(t)}{T(t)} = \frac{q_c}{T_0} \quad ; \quad \tau = \frac{M_\pi}{T(t)} \quad ; \quad \frac{d\tau}{dt} = \tau H(t) \quad ; \quad H(t) = 1.66\, g_{eff}^{\frac{1}{2}}(T)\, \frac{T^2(t)}{M_{Pl}}\,, \tag{10.4.19}$$

furthermore take the overall scale to be M_π and define the following dimensionless ratios

$$m_h \equiv \frac{M_h}{M_\pi} \quad ; \quad m_l \equiv \frac{M_l}{M_\pi} \quad ; \quad l = \mu, e\,. \tag{10.4.20}$$

As argued previously, production from pion decay begins after the QCD crossover during the radiation dominated epoch. We will argue that freeze out occurs at temperatures $T \gtrsim 10MeV$ so that the Hubble factor is given by eq 10.4.19 for the entire period of production. With this change of variables and factoring the constants out of Eq. (10.4.11) leads to the definition of the overall scale of the problem:

$$\Lambda_{lh} = \frac{|H_{lh}|^2}{\sqrt{g(t)}} \left[\frac{|V_{ud}|^2 f_\pi^2(0) G_F^2}{8\pi * 1.66} M_{pl} M_\pi \right] \left(m_l^2 + m_h^2 - (m_l^2 - m_h^2)^2 \right)\,, \tag{10.4.21}$$

where the dimensionless ratios $m_{l,h}$ are defined in (10.4.20). After the temperature falls below the hadronization temperature ($T \sim 155MeV$), the relativistic degrees of freedom ($g(t)$) remain constant until the temperature cools to below the μ mass ($T \sim 106MeV$) and again remains constant until cooling to below the electron mass. For $155MeV \gtrsim T \gtrsim 106MeV$ the relativistic degrees of freedom are $g(t) \sim 14.25$ and for $106MeV \gtrsim T \gtrsim 0.5MeV$ the degrees of freedom are $g(t) \sim 10.75$ [334]. As will be argued, ν_h production is complete by $10MeV$ and since $g(t)$ has a small variation in the temperature range of production, $g(t)$ is

replaced by its average value $\bar{g} = 12.5$. With this, the time variation of the overall scale Λ_h can be neglected, and using $|V_{ud}| = 0.974$[334] we find

$$\Lambda_{lh} \simeq 0.13 \left(\frac{|H_{lh}|^2}{10^{-5}} \right) \left[m_l^2 + m_h^2 - (m_l^2 - m_h^2)^2 \right] . \tag{10.4.22}$$

The function

$$C[m_l, m_h] = \left[m_l^2 + m_h^2 - (m_l^2 - m_h^2)^2 \right] \tag{10.4.23}$$

reveals the usual *chiral suppression* in the case of negligible neutrino masses as a consequence of the (V-A) coupling, being much larger for the μ-channel than the e-channel with $C[m_e, m_h]/C[m_\mu, m_h] \simeq 10^{-4}$ for $m_h \to 0$.

The compilation of bounds on mixing angles of heavy neutrinos in ref.[98] in combination with bounds from X-ray astrophysical data[351, 352] suggests that $|H_{lh}|^2 \ll 10^{-5}$ for few keV $< M_h <$ 140 MeV consequently $\Lambda_{lh} \ll 1$ for M_h in this range.

Production from pion decay begins shortly after the QCD crossover at $T_{QCD} \simeq 155$ MeV when pions form, therefore $M_\pi/T_{QCD} = \tau_0 \leq \tau < \infty$ where $\tau_0 \simeq 0.92$.

Upon substituting the dimensionless variables, Hubble factor and overall scale the kinetic equation may be written in the following form:

$$\frac{1}{\Lambda_{lh}} \frac{dn_{lh}}{d\tau}(y, \tau) = \frac{\tau^2 \left(1 - \frac{M_\pi^2}{6\tau^2 f_\pi^2(0)} \right)}{\sqrt{y^2 + m_h^2 \tau^2} \left(e^{E_h(q,t)/T(t)} + 1 \right)} \ln \left(\frac{e^{E_\pi(p,t)/T(t)} - 1}{e^{-E_h(q,t)/T(t)} e^{E_\pi(p,t)/T(t)} + 1} \right) \Bigg|_{p=p_-(t)}^{p=p_+(t)} . \tag{10.4.24}$$

Upon evaluating the pion energy at the solutions of Eq. (10.7.17), in terms of the dimensionless variables defined above (10.4.19,10.4.20) we obtain[4]

$$\frac{E_\pi(p_\pm(t), t)}{T(t)} = \frac{1}{2m_h^2} \left[\Delta_{lh} \left[y^2 + m_h^2 \tau^2 \right]^{1/2} \pm y \, \delta_{lh} \right] . \tag{10.4.25}$$

where we introduced

$$\Delta_{lh} = 1 - m_l^2 + m_h^2 \tag{10.4.26}$$

$$\delta_{lh} = \left[1 + m_l^4 + m_h^4 - 2m_l^2 - 2m_h^2 - 2m_h^2 m_l^2 \right]^{1/2} = \left[\Delta_{lh}^2 - 4\, m_h^2 \right]^{1/2} \tag{10.4.27}$$

[4]We do not include a label l in the result for E_π to avoid cluttering of notation.

From this equation, the rate can be integrated numerically to study the distribution function after freezeout and obtain the cosmological quantities discussed in section (10.3). However, before we proceed to a numerical integration of (10.4.24) we recognize important features of this equation that anticipate the behavior of the solution:

- The prefactor, proportional to τ^2 is small initially but the argument of the logarithm attains its largest value as the initial temperature $T(t_0) \geq M_\pi$, however, as τ increases, the temperature decreases and the logarithmic term decreases eventually exponentially beating the growth of the prefactor. This analysis indicates that the production rate peaks as a function of time τ and falls off fast. We confirm this behavior numerically in fig. (55). This figure shows the fast rise and eventual fall off of the production *rate* which becomes exponentially suppressed for $\tau \gtrsim 10$. This entails a freeze out of the distribution function: while the rate is exponentially suppressed, the total integral will remain finite, thereby fulfilling the freeze-out condition (10.2.39). In ref. [396] a similar behavior was noticed for lighter neutrinos and we confirmed numerically that a similar pattern holds for the whole kinematic range of M_h in each lepton channel.

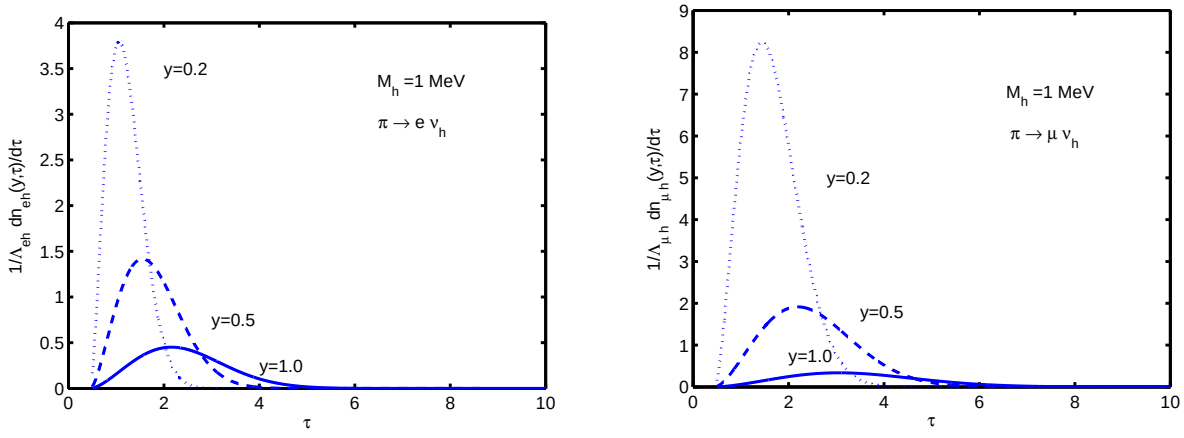


Figure 55: Production rate of a heavy ν_h from $\pi \to l\nu_s$ (eqn. (10.4.24) with $l = \mu, e$ for $M_h = 1MeV$.

- The expression (10.4.25) with (10.4.26,10.4.27) which is a consequence of the kinematics has very important implications on the "kinematic entanglement" of the heavy neutrino.

For fixed y, m_h the ratio (10.4.25) is *smaller* in the μ channel than in the e channel because both Δ_{lh} and δ_{lh} are smaller, this implies that the rate is *larger* in the μ channel. This feature is also displayed in fig. (55). Therefore, we expect that $n_{\mu h}(y,\tau)/\Lambda_{\mu h} > n_{eh}(y,\tau)/\Lambda_{eh}$ and the distributions at freeze out to display this difference. Furthermore, the function $C[m_l, m_h]$ (10.4.23) in Λ_{lh} (10.4.22) is *also* larger for the μ-channel than in the e-channel for small m_h. This is a consequence of the chiral suppression as a consequence of $V-A$ and is displayed in fig.(56).

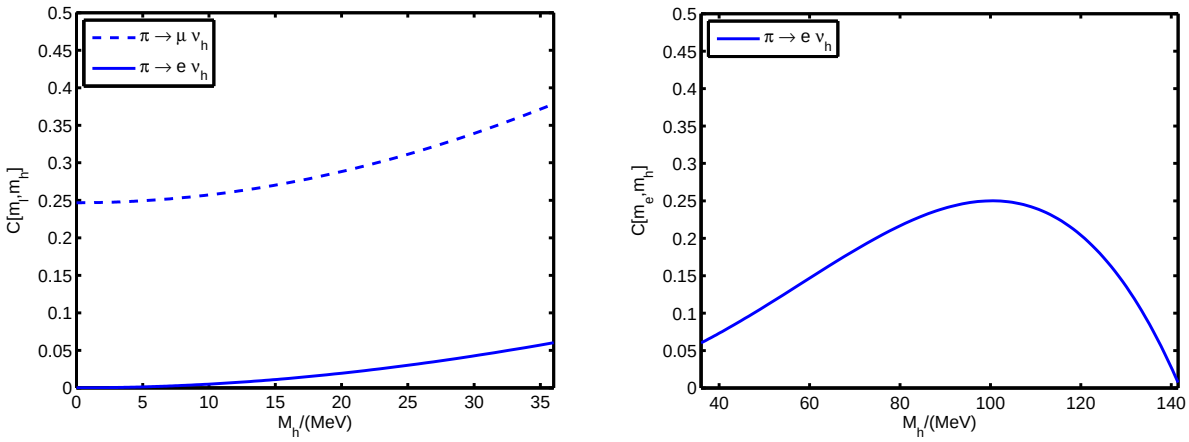


Figure 56: $C[m_\mu, m_h]$ for both channels in the kinematically allowed region of M_h.

The loss rate $\Gamma^{>}$ is obtained from the gain rate (10.4.24) from the detailed balance condition (10.4.5) and we can now proceed to integrate the quantum kinetic equation. However, we first confirm that the gain-loss processes solely from pion decay (and recombination) lead to freeze-out out of LTE, namely we first confirm that the condition (10.2.40) is fulfilled. This is shown explicitly in figs. (57). These figures confirm that the condition (10.2.40) is fulfilled provided $\Lambda_{lh} \ll 1$. The result (10.4.22) indicates that this is the case for $|H_{lh}|^2 \ll 10^{-5}$. Cosmological bounds from X-ray data[351, 352, 393] and a compilation of accelerator bounds[98] suggest that $|H_{lh}|^2 \ll 10^{-5}$ for few keV $\leq M_h \leq 140$ MeV, therefore the condition $\Lambda_{lh} \ll 1$ is satisfied guaranteeing self-consistently that the condition (10.2.40) is fulfilled.

In this case the population build up is obtained by integration of the rate eqn. (10.4.24) as per the general result (10.2.41). In order to exhibit clearly the contribution from π-decay

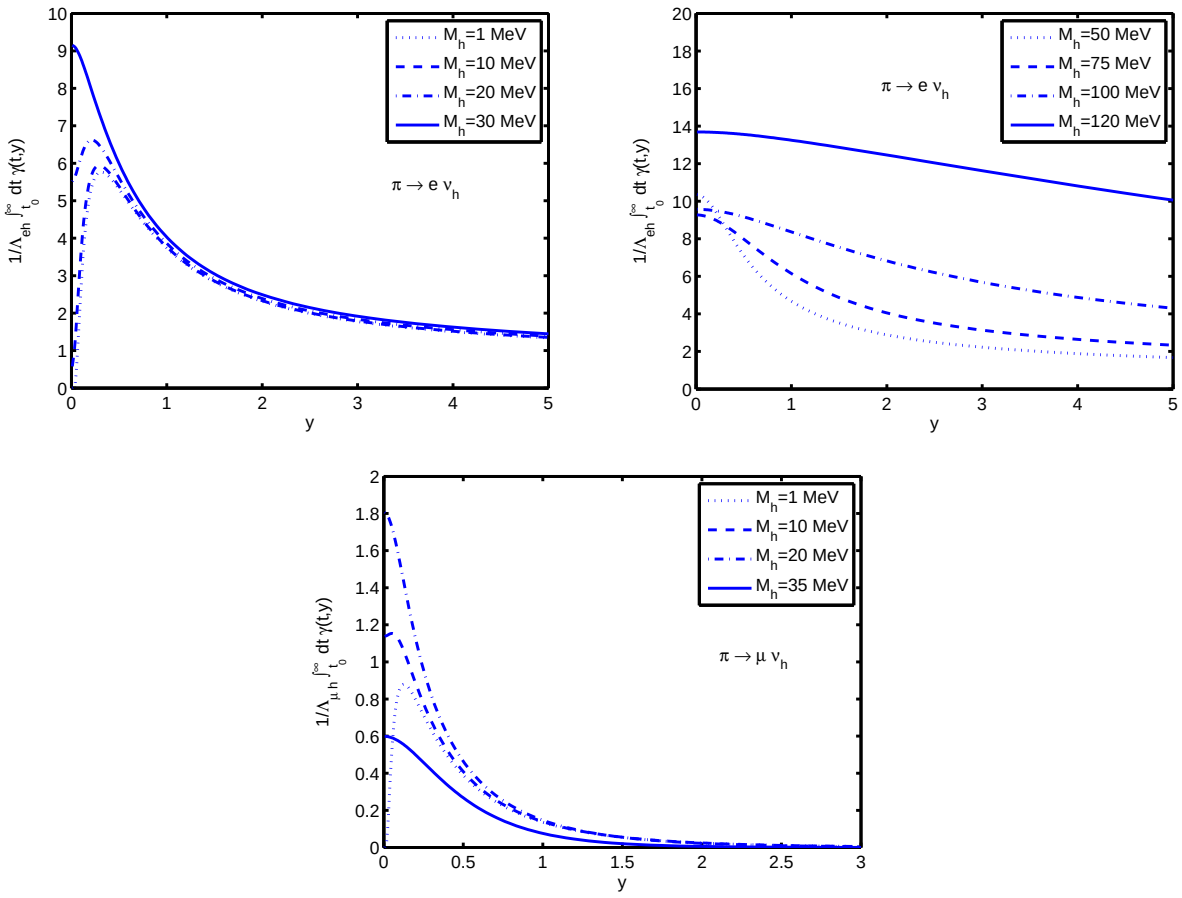


Figure 57: $(1/\Lambda_{lh})\int_{\tau_0}^{\infty}\gamma_h(y,\tau')d\tau'$ for the two channels .

we assume vanishing initial population, in this approximation any initial population must be added. For each channel we obtain the distribution function from direct integration of the gain rate

$$n_{lh}(y,\tau)=\int_{\tau_0}^{\tau}\frac{dn_{lh}}{d\tau}(y,\tau')\,d\tau'\,. \tag{10.4.28}$$

The results are shown in figs (58,59,60). Each figure shows the distribution at different values of time (τ) and, in all cases, the distribution has been frozen out at $\tau\simeq\tau_{fr}\simeq 10$ at which $T(\tau_{fr})\lesssim 14MeV$, this is because the rate is being suppressed through the pion thermal distribution as displayed in figs.(55).

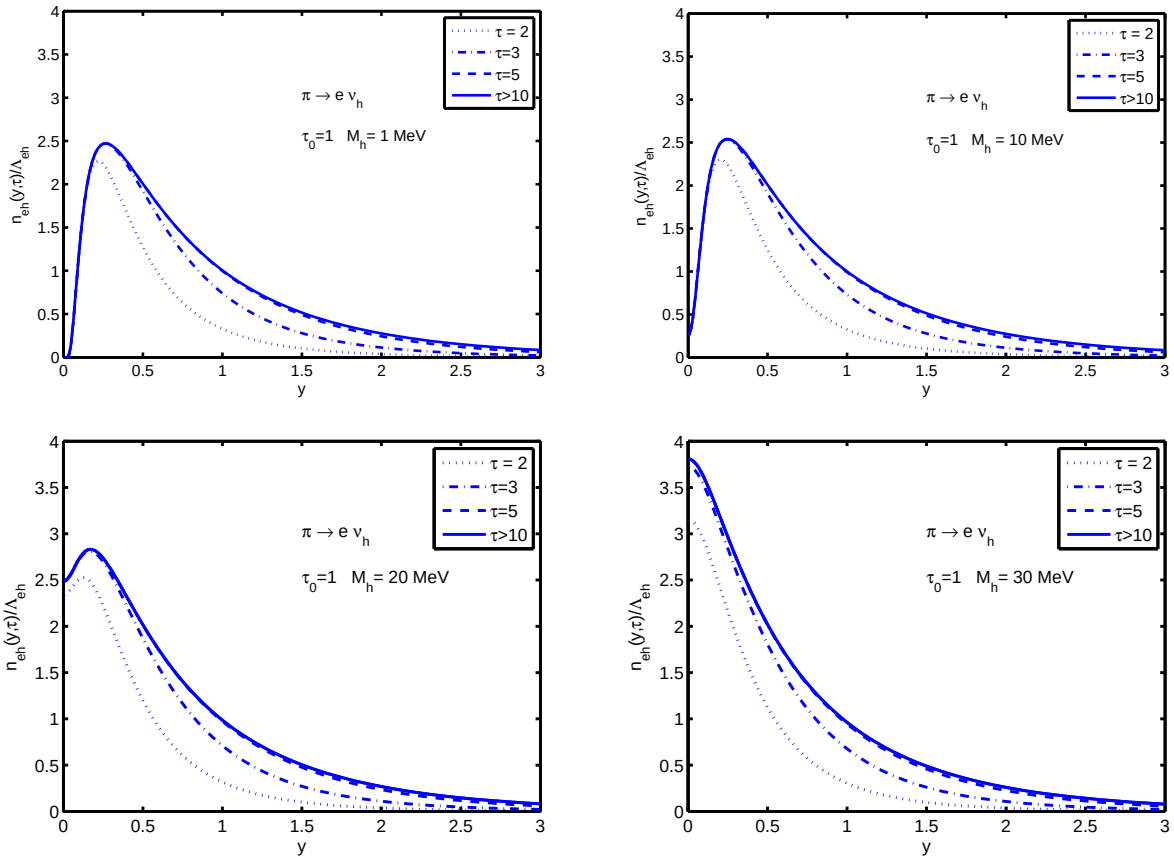


Figure 58: Distribution function $n_{eh}(y,\tau)$ from $\pi \to e\nu_h$ with $M_h = 1, 10, 20, 30MeV$ and vanishing chemical potentials. The solid line is the asymptotic frozen distribution.

There are several features to observe from these plots, the first of which is that, for low mass neutrinos ($M_h \lesssim 1MeV$), the distributions observed in [396] are recovered. For heavier neutrinos ($M_h \gtrsim 30MeV$) a very different behavior from the light species is observed. The light species features a vanishing distribution for small momentum ($y \to 0$), peaks at a particular momentum and falls off at large momentum whereas the heavier species has non-vanishing support at zero momentum ($y = 0$) and monotonically decreases as momentum increases. An intermediate behavior can be observed in the electron channel for a heavy neutrino with $M_h = 20MeV$; the crossover in behavior occurs for $M_h \simeq T(\tau_{fr}) \simeq 10-14\,\mathrm{MeV}$ for $\tau_{fr} \simeq 10$ is the time scale at which the distribution functions freeze out, the freeze out

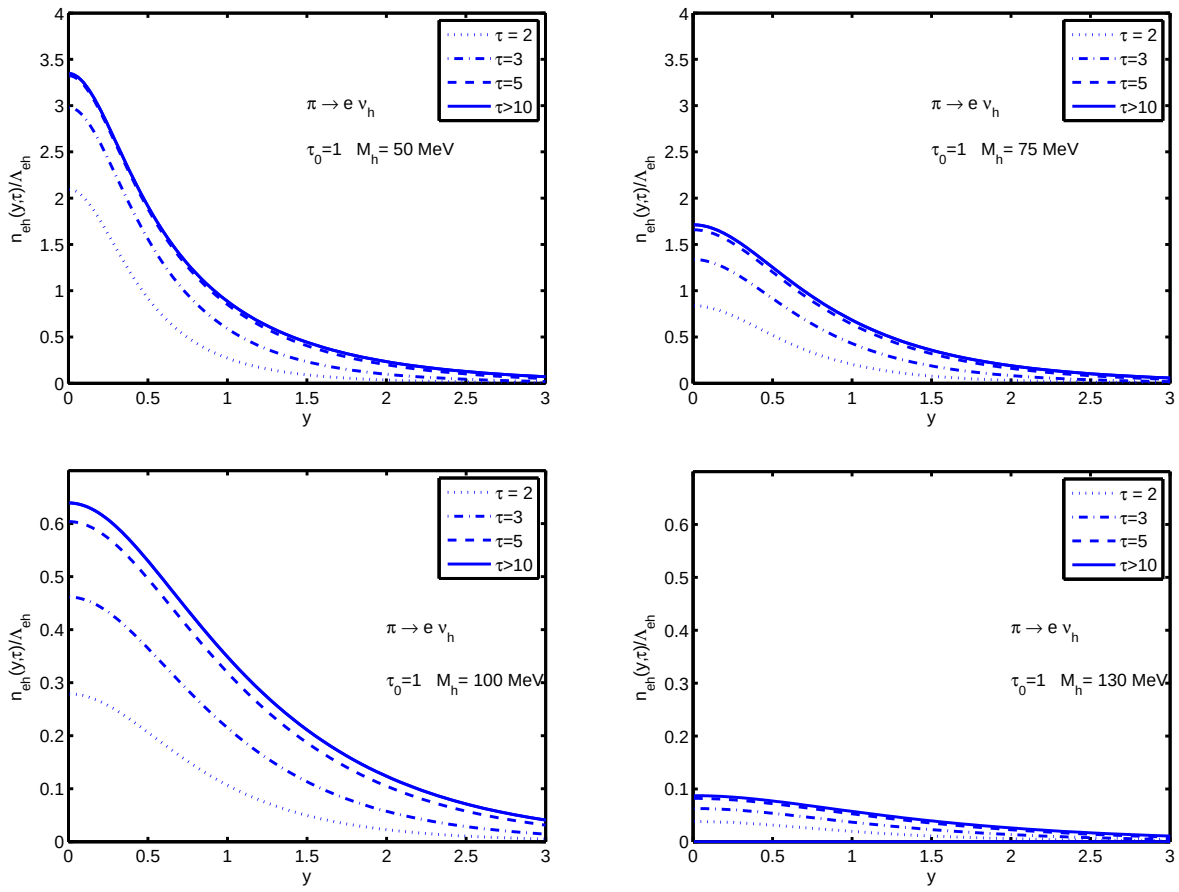


Figure 59: Distribution function $n_{eh}(y,\tau)$ from $\pi \to e\nu_h$ with $M_h = 50, 75, 100, 130 MeV$ and vanishing chemical potentials. The solid line is the asymptotic frozen distribution.

time $\tau \simeq 10$ is nearly independently of the value of M_h.

Furthermore, comparison of figs. (58) with those of figs. (60) is revealing. These two sets display the distributions from the electron and muon channels within the kinematically allowed window available for *both* channels, $0 < M_h \leq 36\,\mathrm{MeV}$, and it is in this comparison that the relevance of the distribution functions n_{lh} for $l = \mu, e$ become manifest. Two important aspects stand out: a) for small y the distribution function from the μ channel is systematically *larger* than that for the e channel, this is a consequence of the kinematic factors Δ_{lh}, δ_{lh} in (10.4.25), both are much smaller in the μ channel than in the e channel,

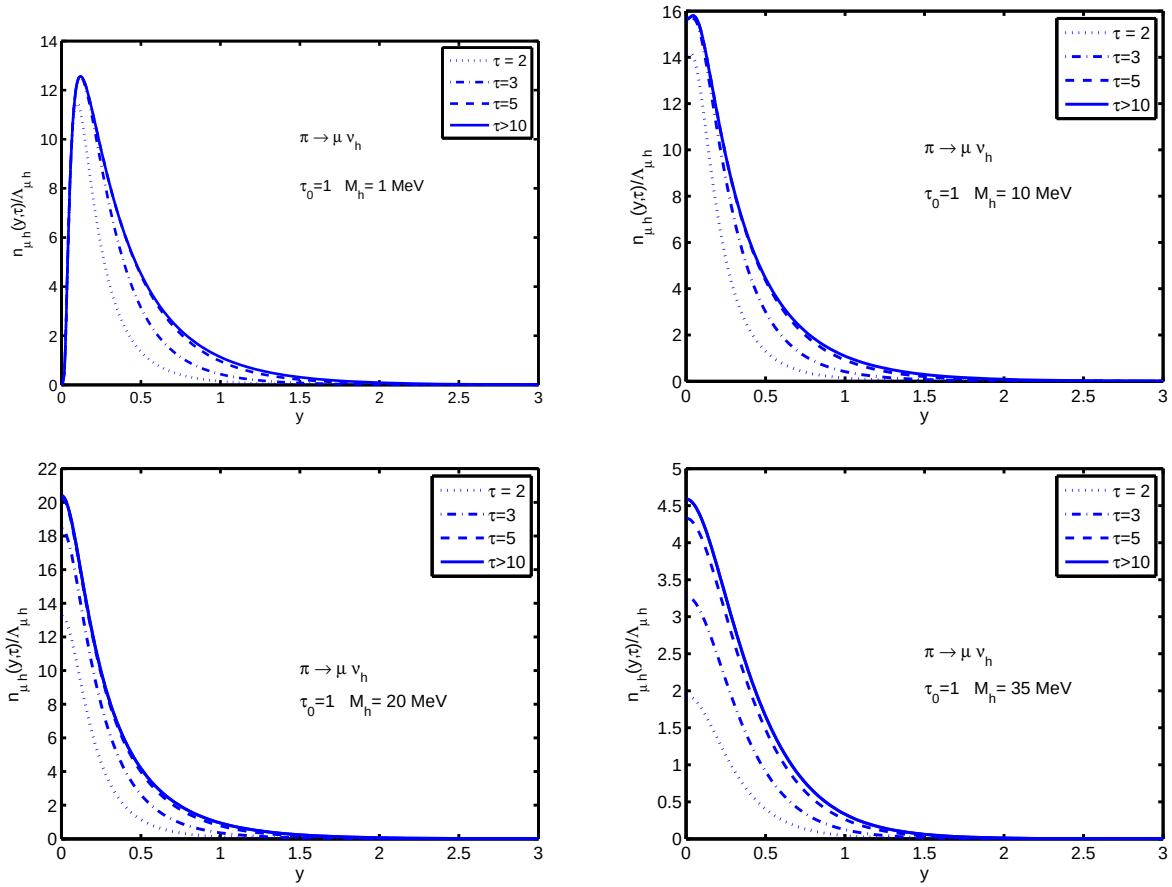


Figure 60: Distribution function $n_{\mu h}(y,\tau)$ from $\pi \to \mu\nu_h$ with $M_h = 1, 10, 20, 35 MeV$ and vanishing chemical potentials. The solid line is the asymptotic frozen distribution.

this fact makes the energies E_π in (10.4.25) *smaller* in the μ channel, therefore yield a *larger* contribution since these are less thermally suppressed. This confirms the analysis presented above. b) The distribution function for the e-channel has larger support for $y > 1$ than that for the μ channel, which is larger for $y < 1$ but us strongly suppressed for $y > 1$ as compared to that of the e-channel. Both these features lead to the conclusion that the distribution function from the μ channel yields a *colder* component than that of the e channel in the sense that the velocity dispersion obtained from $n_{\mu h}$ is smaller than that obtained from n_{eh}, this is a manifestation of the kinematic entanglement and its consequences will be analyzed

in detail below.

The total distribution function is thus a *mixture* of a colder and a warmer component. Defining the distribution function at freeze-out for each channel as

$$f_{lh}(y) = n_{lh}(y, \infty) \tag{10.4.29}$$

where by $\tau = \infty$ we mean $\tau > \tau_{fr} \simeq 10$, the *total* distribution function is given by

$$f_h(y) = \Lambda_{\mu h}\left[\frac{f_{\mu h}(y)}{\Lambda_{\mu h}}\right]\Theta(36\,\mathrm{MeV} - M_h) + \Lambda_{eh}\left[\frac{f_{eh}(y)}{\Lambda_{eh}}\right]\Theta(141\,\mathrm{MeV} - M_h)\,. \tag{10.4.30}$$

The Θ functions describe the thresholds in each channel, the brackets $\left[f_{lh}/\Lambda_{lh}\right]$ are actually the result of the numerical integrations as per eqn. (10.4.24) and are given by the solid lines ($\tau > 10$) in figs. (58-60).

For a given range of masses M_h the only unknowns are the mixing matrix elements H_{lh}. Because the total distribution function is a result of the combination of several production channels, each one with possibly a different mixing matrix element H_{lh}, it is convenient to factor out one of these and rewrite the total distribution function (10.4.30) in terms of ratios, namely

$$f_h(y) \equiv 0.13\left(\frac{|H_{\mu h}|^2}{10^{-5}}\right)\widetilde{f}_h(y) \tag{10.4.31}$$

$$\begin{aligned}\widetilde{f}_h(y) &= C[m_\mu, m_h]\left[\frac{f_{\mu h}(y)}{\Lambda_{\mu h}}\right]\Theta(36\,\mathrm{MeV} - M_h) \\ &+ \left(\frac{|H_{eh}|^2}{|H_{\mu h}|^2}\right)C[m_e, m_h]\left[\frac{f_{eh}(y)}{\Lambda_{eh}}\right]\Theta(141\,\mathrm{MeV} - M_h)\,.\end{aligned} \tag{10.4.32}$$

where the coefficients $C[m_l, m_h]$ are given by (10.4.23) and the brackets $[f/\Lambda]$ are the results obtained numerically and displayed by the solid lines in figs. (58-60). In this manner the various constraints can be phrased in terms of an overall mixing matrix element and ratios, such as $|H_{eh}|^2/|H_{\mu h}|^2$ which along with the kinematic factors $C[m_l, m_h]$ determine the *concentration* of the warmer species (from the electron channel) in the *mixture* of colder and warmer distributions from both channels.

10.4.1 A Tale of Two Distributions:

Closer examination of figs. 58,59,60 reveal a transition in the shapes of distributions as a function of neutrino mass. In the case of relatively light neutrinos with $M_h \lesssim 1 - 10MeV$, the distribution function vanishes at small momentum, reaches a maximum, and falls off with a long tail at high momentum. The situation for heavier neutrinos, $M_h \gtrsim 30MeV$, produces a distribution function which features a non-vanishing plateau at low momentum which steadily falls off with increasing momentum. This latter distribution obviously produces a *colder* dark matter candidate as the distribution function has considerably more support at small momentum. The transition between these two types of distributions is controlled by increasing neutrino mass and an intermediate distribution can be seen in the intermediate mass range, $10MeV \lesssim M_h \lesssim 30MeV$, illustrated by the distribution function in figure 58 for a neutrino mass $M_h = 20MeV$ in the electron channel. For this $M_h = 20MeV$ case, the distribution resembles a superposition of the two distinct distributions seen for the lighter ($M_h \lesssim 10MeV$) and heavier species ($M_h \gtrsim 30MeV$).

Unfortunately, only limited analytic progress can be made towards an understanding of the full distribution function but this proves enough to shed some light on what governs the transition between the different distribution shapes. In reference [396], it is shown that under appropriate approximations, the distribution function for light mass neutrinos takes the form

$$n_{lh}(\tau_0, y)\Big|_{M_h \lesssim 1MeV} = \frac{\Lambda_{lh}}{y^2} \sum_{k=1}^{\infty} \left[\frac{1 + (-1)^{k+1} e^{ky}}{1 + e^y}\right] \frac{\exp\left(-ky/\Delta_{lh}\right)}{k} \times J_k(\tau_0, y) \tag{10.4.33}$$

where

$$J_k(\tau_0, y) = 2\tau_0 \left(\frac{y}{k\Delta_{lh}}\right) \exp\left(-\frac{k\Delta_{lh}\tau_0^2}{4y}\right) + \left(\frac{y}{k\Delta_{lh}}\right)^{1/2} \left[\frac{2y}{k\Delta_{lh}} - \frac{M_\pi^2}{6f_\pi^2}\right] \Gamma\left(1/2, \frac{k\Delta_{lh}\tau_0^2}{4y}\right). \tag{10.4.34}$$

where $\Gamma(\nu, z)$ is the incomplete gamma function and the details of this calculation are reproduced in appendix 10.7.2. As detailed in appendix 10.7.2, the main approximation employed

towards producing this semi-analytic expression is $M_h/M_\pi \ll 1$. This approximation is equivalent to the assumption that the neutrino is produced *ultra-relativistically* ($E_h/T \sim y$) and this proves to be an excellent approximation for lighter species ($M_h \lesssim 1MeV$).

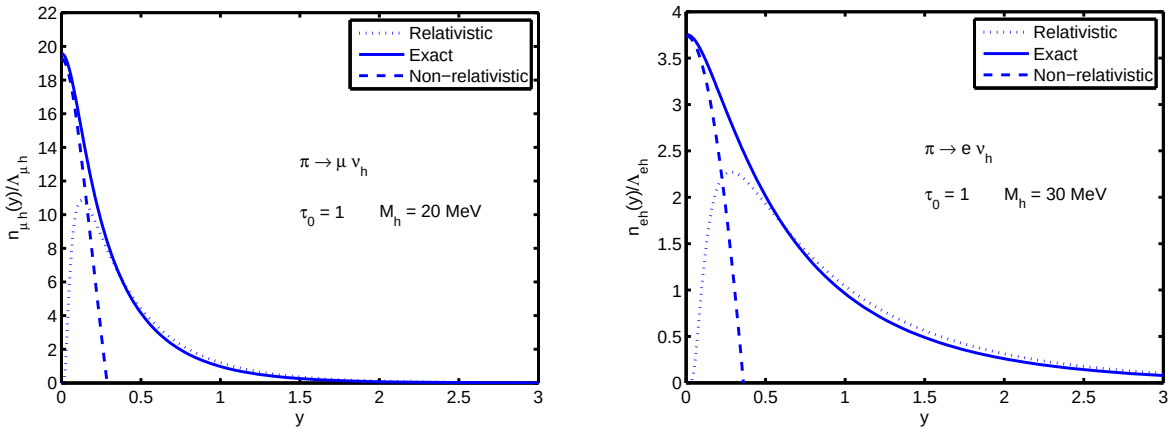


Figure 61: Comparison of the exact distribution with a distribution producing ultra-relativistic and non-relativistic heavy neutrinos.

The opposite type of approximation, where the neutrino is produced *non-relativistically* ($E_h/T \sim M/T + \frac{T}{2M}y^2$), can be made but leads to a very unwieldy expression which is given by Eqs 10.7.23,10.7.30. The production of non-relativistic heavy neutrinos leads to a natural source of non-trivial behavior at small momentum which explains the low momentum plateau. These arguments lend themselves to the interpretation that neutrinos with $M_h \lesssim 10MeV$ are solely produced ultra-relativistically while some fraction are produced non-relativistically as the neutrino mass increases. This argument is illustrated in fig. 61 where the distribution functions of 10.7.28,10.7.30, which are results of ultra-relativistic and non-relativistic production approximations, are plotted against the full distribution. In fig. 61, the low momentum region is dominated by the non-relativistic result whereas the large momentum region fits quite well to the ultra-relativistic result. Both the ultra-relativistic and non-relativistic results fail outside the appropriate momentum regions: the non-relativistic approximation fails for higher momentum whereas the ultra-relativistic approximations fails at decreasing momentum.

This analysis explains the origin of the features of the distribution functions obtained numerically in figures 58, 59, 60, namely that the ultra-relativistic approximation serves as a good estimate for the whole distribution as the neutrino mass becomes negligible compared to the pion mass while the non-relativistic approximation serves to understand the appearance of the plateau at low momentum in the distribution. This means that for light mass neutrinos ($M_h \ll M_\pi$) all of the neutrinos that are produced are done so *ultra-relativistically* while the production of a more massive species leads to some fraction of neutrinos produced *non-relativistically* as well. The fraction of neutrinos which are produced relativistically increases as sterile neutrino mass decreases while the fraction which are produced non-relativistically increases as the sterile neutrino mass increases.

10.4.2 Non-thermality

The equation of state parameter, $w(T)$, is given by eq 10.3.11 where $w = 1/3, 0$ correspond to ultra-relativistic and non-relativistic species respectively. Depending on the temperature of production and decoupling, a heavy neutrino could be ultra-relativistic, non-relativistic ($M \ll T$, $M \gg T$) or somewhere in between as freeze-out occurs and $w(T)$ determines when a particular species becomes non-relativistic.

The correct equation of state (10.3.11) and velocity dispersion (10.3.12) must be obtained from the total distribution function (10.4.30), because these are *moments* of the distribution function they are not simply the addition of the two components.

However, it proves illuminating to *define* an equation of state for each channel

$$w_l(T) = \frac{\mathcal{P}}{\rho} = \frac{1}{3}\frac{\int dy \frac{y^4}{\sqrt{y^2+\frac{M_h^2}{T^2}}} f_{lh}(q_c)}{\int dy\, y^2 \sqrt{y^2+\frac{M_h^2}{T^2}}\, f_{lh}(q_c)}\,, \tag{10.4.35}$$

these serve as *proxies* to quantify the "coldness" of the species produced by the particular channel: from the discussion in section (10.3), $\sqrt{w(T)}$ is a generalization of the "adiabatic speed of sound" for collisionless DM, and in the non-relativistic limit $w(T) \to \langle \vec{V}^2 \rangle/3$.

The distribution function of a thermalized heavy neutrino would be given by the standard

Fermi-Dirac distribution:

$$f_{LTE}(y;T) = \frac{1}{e^{\sqrt{y^2+M_h^2/T^2}}+1}\,, \tag{10.4.36}$$

and we compare the equation of state $w_l(T)$ (10.4.35) with that obtained from (10.4.36), this comparison quantifies the "non-thermality" of the distribution functions from each production channel.

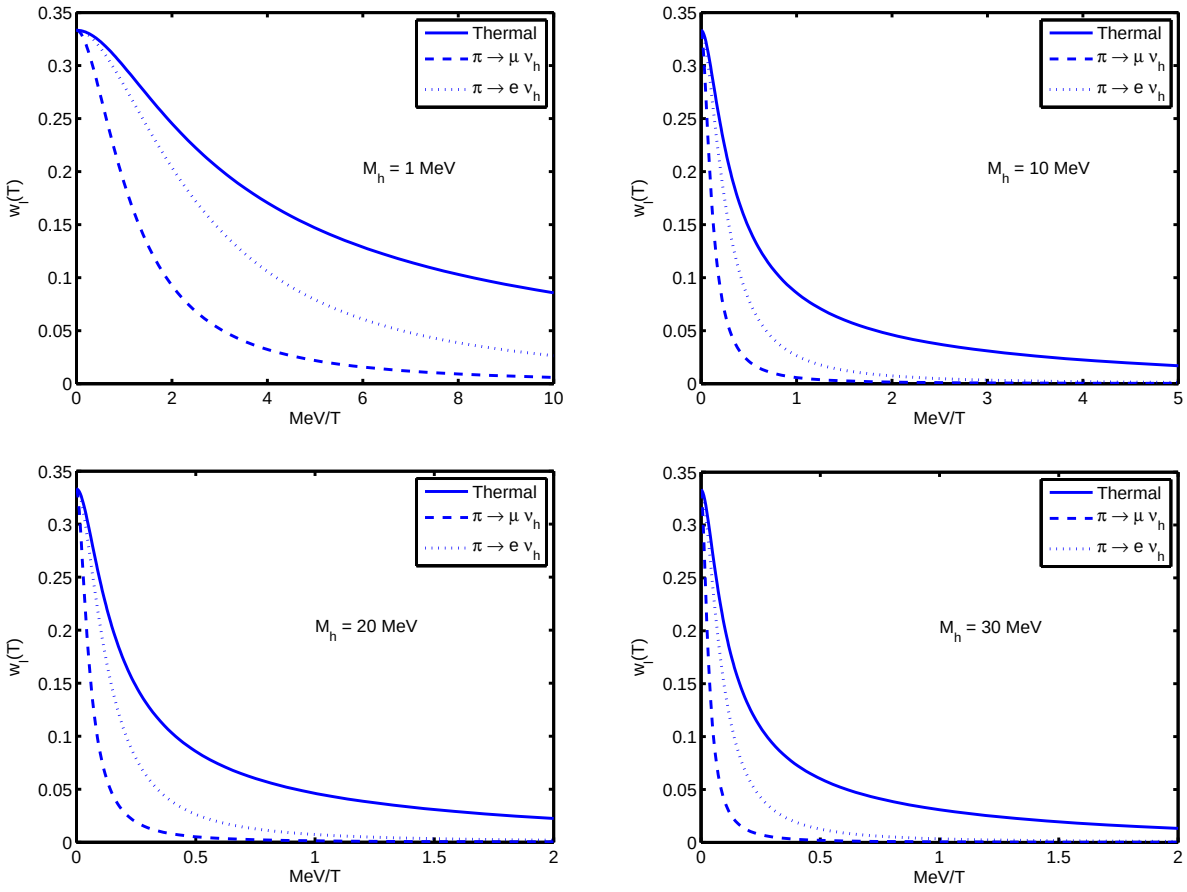


Figure 62: $w_l(T)$ for both channels $\pi \to \mu\nu_h$; $\pi \to e\nu_h$ compared to thermal distribution for $M_h = 1, 10, 20, 30\,\text{MeV}$ in the kinematic window in which both channels are available.

The equation of state $w_l(T)$ for each channel and for the thermal distribution are compared in figs (62,63) as a function of $1/T$ for a range of different masses. These figures clearly display the non-thermality of the heavy neutrino species, furthermore, as anticipated

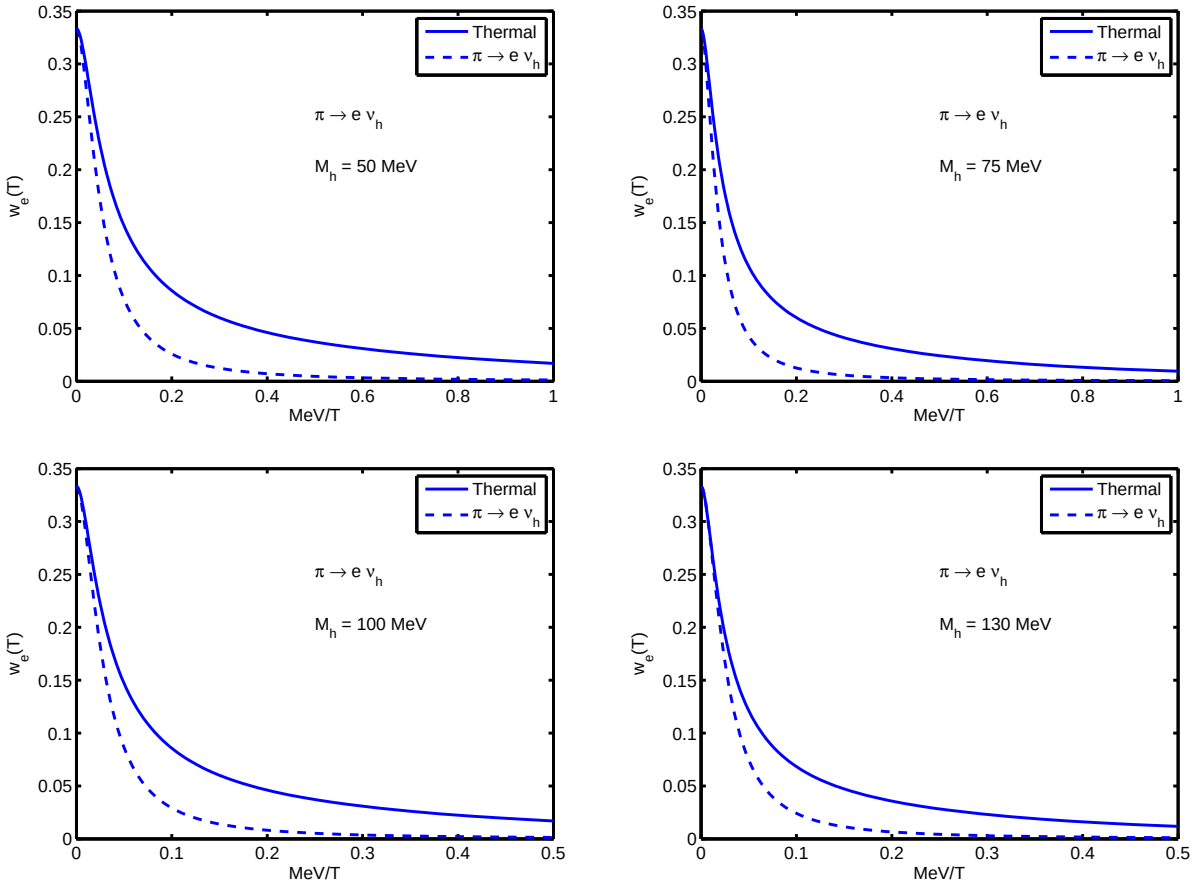


Figure 63: $w_l(T)$ compared to thermal distribution for $M_h = 50, 75, 100, 130\,\mathrm{MeV}$ for $\pi \to e\nu_h$.

by the discussion above, the distribution function from $\pi \to \mu\nu_h$, namely $f_{\mu h}$ yields a *colder* component than that from the e channel, f_{eh}, a direct consequence of the "kinematic entanglement" leading to a larger amplitude at small y for $f_{\mu h}$, and both components are colder and becoming non-relativistic $w(T) \ll 1/3$ much sooner than the thermal case.

The total equation of state is given by the full distribution function (10.4.31,10.4.32), namely

$$w(T) = \frac{\mathcal{P}}{\rho} = \frac{1}{3}\frac{\int dy\, \frac{y^4}{\sqrt{y^2+\frac{M_h^2}{T^2}}}\, \widetilde{f}_h(q_c)}{\int dy\, y^2 \sqrt{y^2+\frac{M_h^2}{T^2}}\, \widetilde{f}_h(q_c)}\,, \tag{10.4.37}$$

which depends on the ratio $|H_{eh}|^2/|H_{lh}|^2$ as this ratio varies between 0 and ∞, it follows that $w(T)$ for $0 < M_h < 36\,\mathrm{MeV}$ interpolates between the two dashed lines corresponding to the muon and electron channels in figs. (62), this is shown in fig. (64). For the mass range $141\,\mathrm{MeV} > M_h > 36\,\mathrm{MeV}$ only the electron channel contributes and $w(T)$ is given by the results displayed in fig.(63).

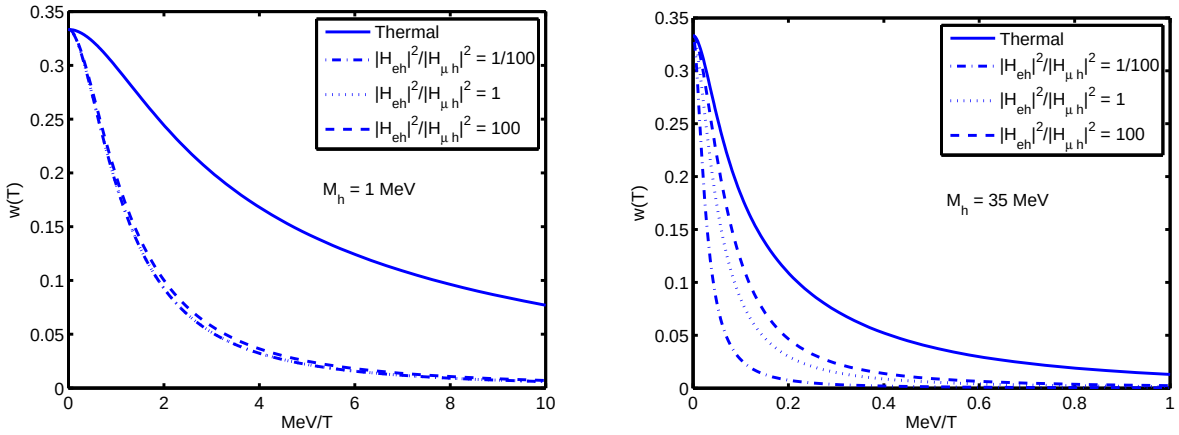


Figure 64: Equation of state with full distribution function for the ratios $|H_{eh}|^2/|H_{\mu h}|^2 = 0.01, 1, 100$ as a function of $(MeV)/T$ for $M_h = 1, 35\,\mathrm{MeV}$. $w(T)$ interpolates between the results for μ and e channels.

10.4.3 Cosmological constraints:

Having obtained the distribution function at freeze-out, we can now implement the general results obtained in section (10.3) and establish the allowed regions within which the various cosmological constraints discussed in section (10.3) are satisfied.

- **Abundance:** with $f_h(y)$ and $\widetilde{f}_h(y)$ given by (10.4.32, 10.4.31) respectively and taking $g_{\nu_h} = 2$ the abundance constraint (10.3.28) becomes

$$2.81\left(\frac{M_h}{\mathrm{keV}}\right)\left(\frac{|H_{\mu h}|^2}{10^{-5}}\right)\int y^2\widetilde{f}_h(y)\,dy \leq 1\,. \tag{10.4.38}$$

This is a function of the ratio $|H_{eh}|^2/|H_{\mu h}|^2$ and also of M_h through the coefficients $C[m_l, m_h]$. This bound was calculated in [396] for $M_h \leq 1MeV$ this mass region is completely dominated by the muon channel, whereas here we allow masses up to $M_h \leq M_\pi - M_l$ which translates to $M_h \leq 142MeV$ for the electron channel and $M_h \leq 36MeV$ for the muon channel.

- **Phase space (Tremaine-Gunn):**

  Using the smallest phase space value of ref [359] which comes from the Fornax dwarf spheroidal galaxy: $(\rho/\sigma^3)_{today} = 2.56\times10^{-4}\,(keV)^4$, and using the general result (10.3.29) with $g_{\nu_h} = 2$ we obtain the constraint

$$\left(\frac{M_h}{\text{keV}}\right)^4 \left(\frac{|H_{\alpha h}|^2}{10^{-5}}\right) \geq 0.0038\, \frac{\left[\int dy\, y^4 \widetilde{f}_h(y)\right]^{3/2}}{\left[\int dy\, y^2 \widetilde{f}_h(y)\right]^{5/2}}\,. \tag{10.4.39}$$

- **Stability:** The "conservative" stability constraint (10.3.32) is independent of the distribution function.

  These bounds and allowed parameter space are shown in fig. (65) along with the parameters of heavy neutrinos from the recently reported X-ray signals[351, 352] ($M_h = 7.1\,\text{keV}, |H_{\alpha h}|^2 = 7\times10^{-11}$). An important observation about this figure is that the inclusion of *both channels* in the total distribution function leads to (marginal) consistency with the claimed X-ray signal. This differs from ref. [396] which claimed that consistency with the X-ray data was supported in the muon channel but not the electron channel. The main differences between these results and those of ref [396] are that we generalize the results to arbitrary mass (as opposed to $\lesssim 1MeV$) and we include *both production channels* as opposed to considering each channel independently.

  An interesting aspect in the two top figures are "kinks" in the abundance and Tremaine-Gunn (phase space density) line, this is a consequence of the kinematic thresholds in the full distribution function (10.4.30).

- **Free streaming scale:** the free streaming wavevector and length scale are given by

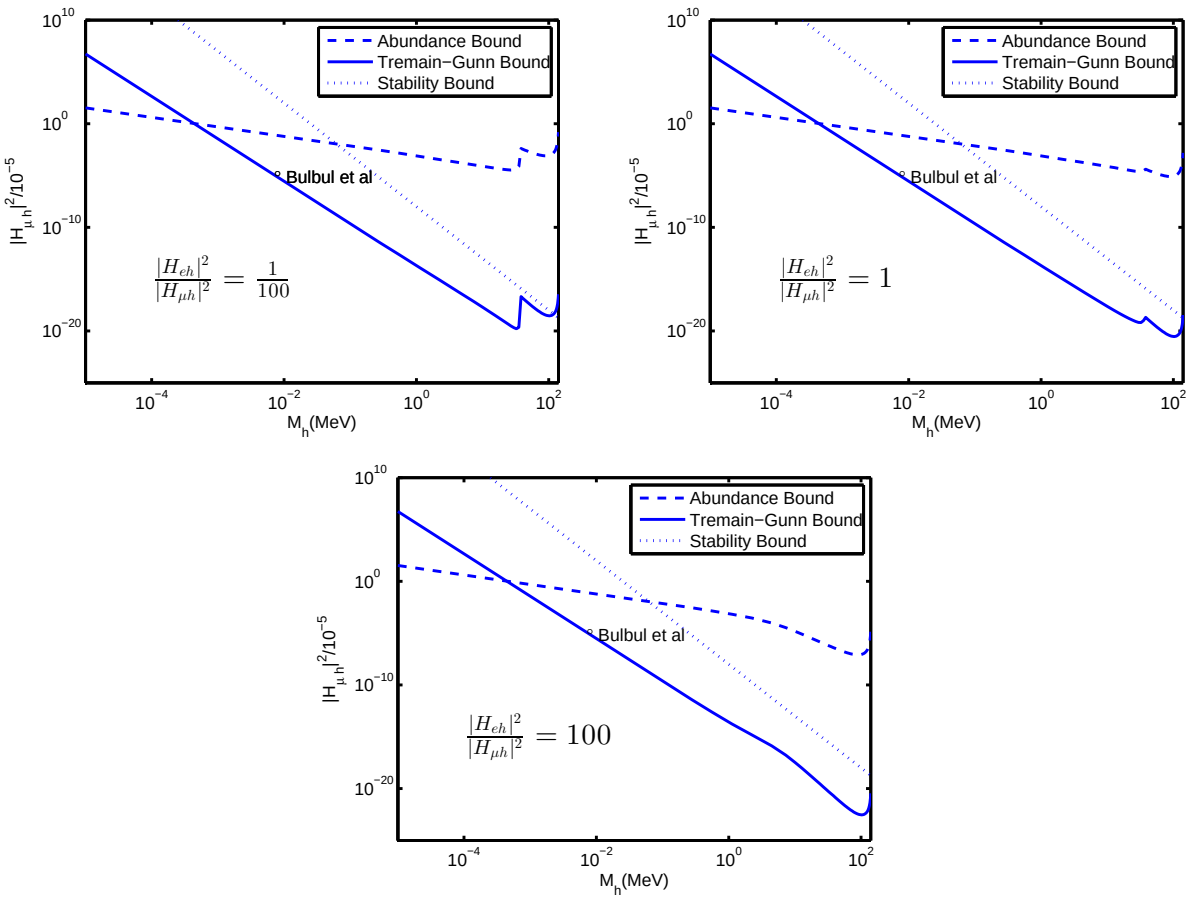


Figure 65: The bounds on $M_h - |H_{lh}|^2$ from abundance, stability and phase space constraints. The allowed regions determined from Eqs (10.4.38,10.4.39 , 10.3.32) are shown and the parameters which potentially explain the 3.5 keV signal[351, 352] are also shown. The kink is a consequence of the thresholds. The allowed parameter space is within the region bound by the three lines.

eqns. (10.3.16-10.3.19) respectively, namely

$$\lambda_{fs,\nu_h}(0) \simeq 5.3 \left(\frac{keV}{M_h}\right) \sqrt{\frac{\int dy\, y^4 \widetilde{f}_h(y)}{\int dy\, y^2 \widetilde{f}_h(y)}}\ \text{kpc}\,. \tag{10.4.40}$$

again this is a non-linear function of the mass and for a species ν_h produced by a single

channel it would be independent of the mixing angle, however if there are several channels, as is the case in π decay, it depends on the ratio of mixing angles. Fig.(66) displays $\lambda_{fs,\nu_h}(0)$ as a function of M_h for various ratios.

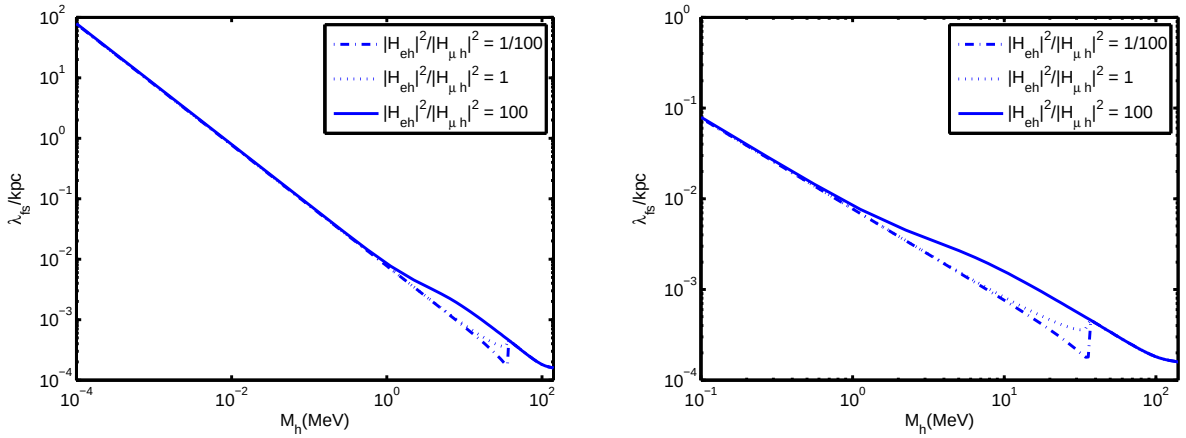


Figure 66: $\lambda_{fs}(0)/(kpc)$ as a function of M_h for the ratios $|H_{eh}|^2/|H_{\mu h}|^2 = 0.01, 1, 100$. Right panel zooms in to highlight the kink because of thresholds and interpolation.

For $M_h \lesssim 1\,\mathrm{MeV}$ the μ (coldest) channel dominates, and $\lambda_{fs}(0)$ is insensitive to the ratio, however, as $M_h \simeq 30\,\mathrm{MeV}$ the contribution of the e channel becomes substantial and dominates above the threshold at $M_h \simeq 36\,\mathrm{MeV}$, the kink in the figure is a result of this threshold. Therefore $\lambda_{fs}(0)$ interpolates between that of the μ and electron channels as a function of the ratio, just as the equation of state interpolates between the colder (μ) and warmer electron dominated components.

10.4.4 Other processes in the same temperature range

In this section we focused on a detailed presentation of heavy neutrino production from pion decay after the QCD hadronization transition, primarily as a clear example where the finite temperature corrections of the pion decay constant and mass have been previously studied in the literature. However, as the discussion in section (10.2) highlights, there are many processes that produce heavy neutrinos via charged and neutral current vertices provided

the kinematics is favorable. In the temperature range just explored $T \simeq 150\,\mathrm{MeV}$ muons are thermally populated and muon decay $\mu \to e\nu_m\nu_h$ is also a production mechanism that is available provided M_h is in the kinematic window for the three body decay. However, this process, is subleading in the temperature range $T \lesssim M_\pi$, this is because the ratio of decay rates

$$\frac{\Gamma_\pi}{\Gamma_\mu} = \frac{\tau_\mu}{\tau_\pi} \simeq 10^2 \tag{10.4.41}$$

however while pions are only available as a production channel below T_{QCD}, muons, on the other hand, are thermally populated at larger temperatures therefore they contribute substantially to the production of heavy neutrinos with masses within the kinematic window. Their contribution to the total abundance of heavy neutrinos merits a deeper study, along with all the other mechanisms described in section (10.2). Furthermore, light neutrinos are thermally populated in a much wider temperature range and $\nu_{m1}\nu_{m2}\nu_{m3} \to \nu_h$ is the three body "fusion" process described by the gain term (10.2.21) that yields heavy neutrinos at temperatures $T \geq M_h$. The inverse process, $\nu_h \to \nu_{m1}\nu_{m2}\nu_{m3}$ is the decay process described by the loss term (10.2.22), which contributes even at zero temperature and describes the decay of the heavy neutrino (10.2.55). The detailed study of all of these processes clearly defines an extensive program to assess reliably the production of heavy neutrinos in cosmology.

10.4.5 Comparison with Dodelson-Widrow[69]:

Although the Dodelson-Widrow (DW)[69] (non-resonant) mechanism of sterile neutrino production via active-sterile oscillations has been recently shown to be inconsistent with cosmological data at $> 99\%$ confidence level[357], it is illustrative to compare the results obtained above for the non-equilibrium distribution function to the (DW) case for further understanding of its cosmological consequences.

The (DW) distribution function is

$$f_{DW}(y) = \frac{\beta}{e^y + 1}\,, \tag{10.4.42}$$

where β is determined by saturating the DM abundance, namely[69, 357]

$$\beta = \Omega_{DM} h^2 \left(\frac{94\, eV}{M_h}\right) = \frac{11.3\, eV}{M_h}\,. \tag{10.4.43}$$

Furthermore, we will focus our comparison on the mass scale $M_h \simeq 7\, keV$ because this scale is of observational relevance[351, 352] and we consider the μ channel since for $M_h \simeq 7\, keV$ this channel features the largest branching ratio. The distribution function obtained from pion decay, $f_h(y)$ is defined by eqn. (10.4.29). An important ingredient in the comparison are the following results:

$$\frac{1}{\beta}\int dy\, y^2 f_{DW}(y) \;=\; 1.803 \tag{10.4.44}$$

$$\frac{1}{\Lambda_{\mu h}}\int dy\, y^2 f_h(y) \;=\; 1.830 \tag{10.4.45}$$

therefore the total integral of the distribution functions *divided by their prefactors* is *approximately the same*, however as gleaned from fig.60 $f_h(y)/\Lambda_{\mu h}$ is strongly peaked at small momenta, and as discussed in the text the distribution function is fairly insensitive to M_h for $M_h \lesssim 1\, MeV$. We provide two different manners to compare the distributions; i) fixing β and $\Lambda_{\mu h}$ to saturate the DM density in both cases ii) for fixed values of M_h and mixing angles $|H_{\mu h}|^2$ extracted from the analysis of ref.[357] to obtain $\Lambda_{\mu h}$ but consistent with the (DW) scenario we keep the value of β that saturates Ω_{DM}.

Comparison 1: both β and $\Lambda_{\mu h}$ are fixed to yield the total DM density, namely the fraction (10.3.10) is $\mathcal{F} = 1$ in both cases. This yields

$$\beta = 1.611 \times 10^{-3} \quad , \quad \Lambda_{\mu h} = 1.600 \times 10^{-3} \tag{10.4.46}$$

with these values we display both distribution functions in fig.67.

Although the integrals of the distribution functions are the same, obviously $f_h(y)$ is sharply localized at smaller momenta, therefore yielding a colder distribution.

Comparison2: for this case we keep $\beta = 1.61 \times 10^{-3}$ so that f_{DW} saturates the abundance bound, but $\Lambda_{\mu h}$ is now extracted from the upper bounds on the confidence band for

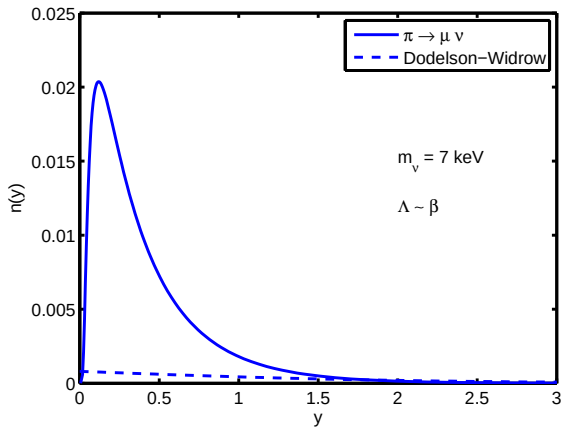


Figure 67: The distribution functions $f_{DW}(y)$ (dashed line) and $f_h(y)$ solid line, for $\beta \simeq \Lambda$. In this case both distribution functions are fixed to saturate the DM density. $M_h = 7\, keV$

the (DW) species of fig. (4) in ref.[357], identifying the mixing matrix element $|H_{\mu h}|^2$ with $\sin^2(2\theta)$ in this reference. We read from this fig. the values

$$M_h \simeq 7\, keV \quad ; \quad |H_{\mu h}|^2 \simeq 10^{-7} \tag{10.4.47}$$

with these values we find

$$\Lambda_{\mu h} = 3.18 \times 10^{-4} \tag{10.4.48}$$

the comparison between the distribution functions is displayed in fig.68. In this case the distribution function $f_h(y)$ yields a fraction $\mathcal{F} = 0.205$ to the DM abundance.

Again it is clear that even when the distribution function yields a smaller fraction, it features a larger contribution at smaller momenta and falls off faster at larger momenta, this feature makes this species *colder* than (DW) with an abundance that while smaller than (DW) is fairly substantial.

A further illuminating comparison is obtained from the free streaming length (10.3.19), which is *independent* of the normalization factors $\beta, \Lambda_{\mu h}$ respectively. We find

$$\lambda_{fs}^{(DW)}(0) = 2.7116\, kpc \tag{10.4.49}$$

$$\lambda_{fs}^{h}(0) = 1.0231\, kpc\,, \tag{10.4.50}$$

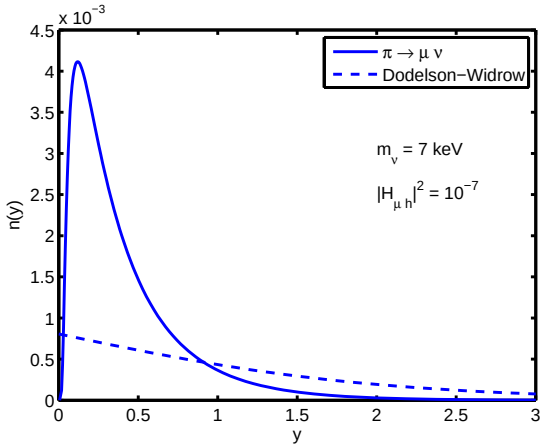


Figure 68: The distribution functions $f_{DW}(y)$ (dashed line) and $f_h(y)$ solid line, for $\beta = 1.61 \times 10^{-3}$; $\Lambda_{\mu h} = 3.18 \times 10^{-4}$. In this case the (DW) distribution function is fixed to saturate the DM abundance, whereas $\Lambda_{\mu h}; M_h = 7\, keV$ are fixed by the upper limits of the confidence band for (DW) in fig. (4) of ref.[357]. For this case f_h yields a fraction $\mathcal{F} = 0.205$ of the DM density.

the low momentum enhancement of $f_h(y)$ yields a much colder distribution with a much shorter free streaming length as compared to the (DW) case.

We conclude that while the (DW) mechanism seems to be ruled out as a *sole* production channel of sterile neutrinos, the comparison with the non-equilibrium function obtained from pion decay in the dominant channel offers a useful insight into the properties of sterile neutrinos produced via this mechanism and suggests that the abundance from this alternative scenario could be a substantial contribution to the DM component.

10.4.6 Comments:

We have implemented the results on the finite temperature dependence of the pion decay constant and pion mass from a substantial body of work on chiral perturbation theory including resonances and linear and non-linear sigma models including contributions from vector

mesons[373, 372, 377, 374, 376]. Taken together this body of results offer a consistent description of the temperature dependence. However, we recognize that there remain uncertainties inherent in an effective description of hadronic degrees of freedom but questioning the validity of ChPT is beyond the scope of this work. Recently a lattice study[429] reported results some of which are at odds with those of refs.[373, 372, 377, 374, 376] such as an increase in the pion decay constant and a decrease of pion temperature near the QCD crossover scale. While a confirmation or rebuttal of these results and/or a resolution of the controversy is awaiting, we can speculate on the impact of the results of this reference, if these hold up. First: if the pion decay constant is *larger* this obviously results in a larger production rate, secondly: if the pion mass is smaller, this also leads to a larger production rate (less thermal suppression of pions in the medium) and *crucially* to a delayed freeze out with a longer stage of production as pions remain populated in the medium for a longer time. Lastly: *if* the pion mass falls below the muon channel threshold at high temperature, the production is solely through the electron channel resulting in a *warmer* distribution, as the pion mass increases towards smaller temperature, the muon channel opens up and the muon production channel with a *colder* component becomes available thus confirming the argument on the "mixed" contributions to the distribution function. All of these aspects *bolster* the case for consideration of pions as an important production mechanism, and the last point in particular, bolsters the argument on kinematic entanglement. We thus conclude that the features imprinted on the non-equilibrium distribution function from the various decay channels and *mixed* nature of the distribution function is a robust qualitative prediction.

10.5 UNSTABLE HEAVY NEUTRINOS: CASCADE DECAYS INTO STABLE DM

Unstable heavy neutrinos with lifetimes much smaller than $1/H_0$ do not play a *direct* role as a viable DM candidate, however they can decay via a *cascade* into lighter stable heavy neutrinos that could be viable candidates. To discuss this scenario in more concrete terms, let us consider a hierarchy of two heavy neutrinos ν_{h_1}, ν_{h_2} with $M_{h_1} \gg M_{h_2}$ and assume that

the heavier, ν_{h_1} is in the kinematically allowed window that allows its production on-shell from π decay, namely $\pi \to l\nu_{h_1}$. If so after being produced, ν_{h1} will cascade decay into ν_{h_2} + leptons on a time scale $\simeq \tau_{\nu_{h1}}$, namely the intermediate heavier ν_{h_1} yields yet another production channel to the lighter ν_{h_2},

$$\pi \to l\nu_{h_1} \to l\nu_{h_2} l_1 l_2 \,, \tag{10.5.1}$$

where l_1, l_2 are other charged or neutral leptons. In this process the intermediate ν_{h_1} goes on shell and the cascade is mediated by resonant decay. In this scenario the production rate of the stable(r) species ν_{h_2} is given by[431, 405, 400]

$$\Gamma^{<}_{\pi\to\nu_{h_2} l l_1 l_2} = \Gamma^{<}_{\pi\to\nu_{h_1} l} \times \mathrm{Br}(\nu_{h_1} \to \nu_{h_2} l_1 l_2) \,, \tag{10.5.2}$$

where $\mathrm{Br}(\nu_{h_1} \to \nu_{h_2} l_1 l_2)$ is the branching ratio. If ν_{h_1} decays into a lighter heavy ν_{h_2} it can also decay into the active-like- light neutrinos ν_m, the decay amplitude for $\nu_{h1} \to \nu_{m_1}\nu_{m_2}\nu_{m_3} \propto HU$, whereas the amplitude to decay into $\nu_{h_1} \to \nu_{h_2} l_1 l_2 \propto H^2$ therefore $Br \propto H^2$ and $\Gamma^{<}_{\pi\to\nu_{h_2} l l_1 l_2} \propto H^4$ namely is suppressed by an extra factor $|H|^2$. If the lifetime of ν_{h_1} is $\gtrsim 10^3\, secs$ it can decay into active-like neutrinos well after Big Bang Nucleosynthesis (BBN) avoiding the constraints on the number of relativistic active neutrinos during BBN providing a late injection of neutrinos into the cosmic neutrino background well after BBN. A lifetime $\gtrsim 10^{11}$ secs would inject a lighter heavy neutrino as a DM candidate after matter radiation equality, just when density perturbations begin to grow under gravitational collapse. Two specific examples illustrate these possibilities: a) consider $M_h \lesssim 1\,\mathrm{MeV}$ from the discussion in section (10.2.5) the decay channel with largest branching ratio ($\simeq 99\%$) is the "invisible" channel $\nu_h \to 3\nu_m$ (see eqn. (10.2.55)) with a lifetime

$$\tau \simeq \frac{10^5}{|H_{mh}|^2} \left(\frac{\mathrm{MeV}}{M_h}\right)^5 s \,, \tag{10.5.3}$$

with $M_h \lesssim 1\,\mathrm{MeV}$; $|H_{mh}|^2 \lesssim 10^{-6}$ the heavy neutrino decays into active-like neutrinos after matter-radiation equality "injecting" a non-LTE component in the cosmic neutrino background after neutrino decoupling. b) The decay rate into a lighter ν_{h_2} is suppressed by another power of $|H_{mh}|^2$, therefore a ν_{h_1} with $M_{h_1} \lesssim 10\,\mathrm{MeV}$; $|H_{mh}|^2 \lesssim 10^{-6}$ can decay into

ν_{h_2} with $M_{h_1} \simeq$ few keV also after matter-radiation equality, now providing a heavy neutrino with a keV mass range as a DM candidate just at the time when density perturbations begin to grow. This latter case may yield an excess of positrons, however, to assess whether this is observationally significant requires a deeper study.

Obviously these are *conjectures* that merit a far deeper analysis, however this scenario is similar to that posited in ref.[411] of a hierarchy of heavy degrees of freedom decaying in a cascade into lighter species that may act as stable(r) DM candidates.

10.6 SUMMARY, DISCUSSION AND FURTHER QUESTIONS

The main premise of our study is that if sterile neutrinos are a suitable extension beyond the Standard Model in which these mix with active neutrinos via an off-diagonal mass matrix, diagonalization of the mass matrix to the mass basis implies that heavy neutrinos mass eigenstates couple to standard model leptons via charged and neutral current interactions. The same processes that produce active-like light neutrinos also produce heavier neutrinos if kinematically allowed, albeit with a much smaller branching ratio determined by small mixing angles. We study the production of heavy neutrinos *via standard model charged and neutral current interactions* under a minimal set of assumptions: small mixing angles with flavor neutrinos, standard model particles are in local thermodynamic equilibrium. We obtain the quantum kinetic equations that describe their production to leading order in the small mixing angles and give the general solution in terms of gain and loss rates that obey detailed balance. A wide range of charged and neutral current processes available throughout the thermal history of the Universe lead to cosmological production of heavy neutrinos including the possibility of production from *collective excitations* and "rare" processes in the medium such as plasmon decay, and "inverse" processes such as $\gamma\nu_m \to \nu_h$.

We discuss the general conditions for thermalization and argue that heavy neutrinos with lifetimes $> 1/H_0$ (the Hubble time scale) freeze-out with non-equilibrium distribution functions. We generalize the concept of *mixed DM* to the case in which a single species of

heavy neutrinos is produced by different channels and argue that in each channel the heavy neutrinos produced are *kinematically entangled* with the lepton produced in the reaction. If the distribution function freezes out of local thermal equilibrium it maintains memory of this kinematic entanglement in the form of a colder or warmer distribution function as compared to other channels. We quantify the "coldness" by obtaining the equation of state parameter for each channel, which serves as a "proxy" for the velocity dispersion of the DM particle when it becomes non-relativistic. If several channels contribute to the production of a particular species, the total distribution function is a *mixture* with components produced from the different channels, these may be colder or warmer as a consequence of the kinematic entanglement. The concentration of each component depends on the kinematics and the ratio of mixing angles in each channel.

We summarize the abundance, phase space density and *stability* constraints that a suitable DM candidate must fulfill in terms of the total distribution function at freeze-out and discuss clustering properties such as primordial velocity dispersions and free streaming lengths. We compared the quantum kinetic framework with other treatments available in the literature, recognizing the necessity for a consistent non-perturbative treatment in the case of possible MSW resonances in the medium.

An important conclusion of this analysis is that, whereas many efforts focus on some temperature scale and on particular production processes, in order to reliably assess the feasibility of a particular heavy neutrino DM candidate, *all* possible production channels of this species must be carefully analyzed throughout the thermal history of the Universe. An immediate consequence of this principle is to alter the "standard" production mechanisms - Dodelson-Widrow, Shi-Fuller, scalar decays - all of which assume a vanishing initial population. The example of pion decay provides a unambiguous source of sterile neutrino population which will modify the standard results in a prescription described in [398]. Further work is in progress on this front.

We argued that the final distribution function of heavy neutrinos after freeze-out is indeed a *mixture* of the various contributions to it from the different production processes that are kinematically available. Only in the case of a fully thermalized population will the

"memory" from the different processes be erased, but non-equilibrium distribution functions will have imprinted in them the kinematic entanglement from the different processes.

As an explicit example we studied the production of heavy neutrinos from charged pion decay after the QCD crossover into the hadronized phase within the effective field theory of weak interactions of charged pions, including finite temperature corrections to the pion decay constant and mass. Pion decay is one of the main sources of neutrino beams in accelerator experiments and pions being the lightest hadrons formed after the QCD crossover, their decay surely contributed to the cosmological production of heavy neutrinos. While not claiming that this process is more or less important than others (a claim that requires a detailed assessment of the other production mechanisms) it provides a wide kinematic window $M_h \lesssim 140\,\mathrm{MeV}$ from two different channels and offers a clear example of kinematic entanglement: the distribution function from the μ channel is distinctly *colder* than that of the electron channel and the total distribution is a mixture of both. We obtain the allowed region of parameters that fulfill the abundance, phase space density and stability constraints, these are displayed in figs. (65) for various values of the ratio $|H_{eh}|^2/|H_{\mu h}|^2$, the boundaries of these regions reveal the thresholds for the different channels, a hallmark of kinematic entanglement. The equation of state (or alternatively velocity dispersion) and free streaming length interpolate between the colder component and the warmer component from the μ and e channels respectively, as a function of M_h and the *ratio* $|H_{eh}|^2/|H_{\mu h}|^2$, clearly displaying the *mixed* nature of the total distribution function.

We *conjecture* that heavy neutrinos with lifetimes $\ll 1/H_0$ may decay into active-like neutrinos after neutrino decoupling injecting light neutrinos out of equilibrium into the cosmic neutrino background, and for $M_h \lesssim 10\,\mathrm{MeV}$ and $|H_{lh}|^2 \lesssim 10^{-6}$ *may* decay after matter radiation equality into another heavy but lighter and stabler neutrino that may be a suitable DM candidate.

Furthermore, we have provided a detailed comparison between the non-equilibrium (non-thermal) distribution function obtained from pion decay and that of the Dodelson-Widrow scenario. The comparison was carried out within two different scenarios: i) saturation of DM abundance for each case, and ii) the parameters $M_h, |H_{lh}|$ were extracted from the

upper bounds of the confidence band in the analysis of ref.[357]. In the second case the distribution function from pion decay yields a substantial fraction of DM, in both cases the distribution function from pion decay yields much shorter free streaming lengths. This comparison suggests that production via pion decay *could* yield a substantial contribution to the DM abundance with a species that is colder than a thermal species but without invoking resonant production via a leptonic asymmetry.

The wide range of production mechanisms available throughout the thermal history of the Universe explored in this study suggests the necessity of an exhaustive program to assess their contributions to DM from heavy neutrinos.

Further Questions.

Our study raises important questions that merit further and deeper investigation: **i:)** we recognize that the possibility of MSW resonances in the medium would require going beyond the leading order in the mixing matrix element to obtain the quantum kinetic equations (this is also a caveat of the approach in ref.[390]), one possible avenue would be to obtain the non-equilibrium effective action for the neutrino sector by tracing over the remaining degrees of freedom of the standard model (quarks, charged leptons, vector bosons) assumed to be in LTE. **ii:)** we argued that *collective excitations in the medium* could lead to the production of heavy neutrinos at high temperature, for example via plasmon decay, this is a cooling mechanism of stars in advanced stages of stellar evolution, and its high temperature counterpart in the early Universe *may* be a suitable production mechanism of heavy neutrinos. This is an intriguing possibility that requires a thorough analysis implementing the hard thermal loop program[418, 419, 421] to obtain the plasmon dispersion relations and couplings to neutrinos. **iii:)** if there is a hierarchy of heavy neutrinos there is the possibility of mixed DM: the contribution from different heavy neutrinos providing cold and warm components depending on their masses and distribution functions. In this case (unlike the case where mixed DM arises solely from one component but from various production channels) it is far from clear what is the *effective free streaming length and coarse grained phase space density*. Since the free streaming length determines the cutoff scale in the linear power spectrum of density perturbations it is important to obtain a reliable assessment of

its interpretation in the case of mixed DM, is it possible that a very heavy neutrino (cold DM) and a lighter one (hot DM) can mimic warm DM?, is it possible that such mixture would lead to the same power spectrum as one single species of mass $\simeq 7\,\mathrm{keV}$?. In order to shed light on these questions, the collisionless Boltzmann equation for several DM components in an expanding cosmology must be studied. Similar questions apply to the *effective* coarse grained phase space density, as this quantity is observationally accessible (or inferred) from the kinematics of dwarf spheroidal galaxies and provides a powerful constraint on a DM candidate independently of abundance. We expect to report on some answers to these questions in later studies.

10.7 APPENDICES

10.7.1 Quantum Kinetic Equation

In this appendix we set up the quantum kinetic equation describing the sterile neutrino population arising from the reaction $\pi^{\pm} \rightleftharpoons l^{\pm}\nu_s(\bar{\nu}_s)$. The kinetics were originally set up in [396] but is repeated here for completeness. This process will occur when the plasma is at temperatures below the QCD phase transition and it is acceptable to use an effective field theory which treat pions as the fundamental degrees of freedom. The low-energy effective interaction Hamiltonian responsible for this process is given by eqn. (10.4.4) including finite temperature corrections to f_π.

The population buildup of the heavy neutrinos is described with a quantum kinetic equation given by form (10.2.8) where the gain and loss terms are calculated by writing the quantum mechanical transition amplitudes from an initial state i to a final state f, $|\mathcal{A}_{fi}|^2$, and integrating over kinematic region of phase space. With the effective Hamiltonian (10.4.4), the relevant reactions are displayed in fig 69.

The gain terms are due to the decay process $\pi^+ \to \bar{l}\nu_l$ which starts from an initial state with $n_\pi(\vec{p})$ pion quanta and $n_\alpha(\vec{k}), n_h(\vec{q})$ charged lepton and heavy neutrino mass eigenstate respectively while the final state has $n_\pi(\vec{p})-1, n_\alpha(\vec{k})+1, n_h(\vec{q})+1$ quanta for each respective

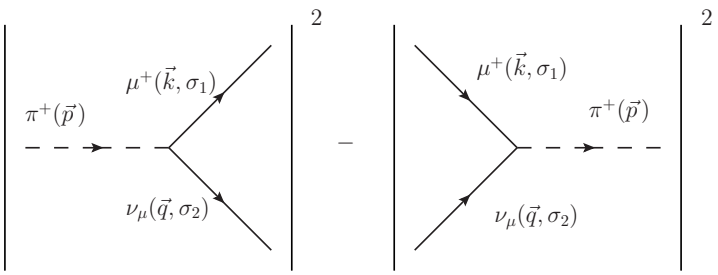


Figure 69: The gain/loss terms for the quantum kinetic equation describing $\pi^+ \to \bar{\mu}\nu_\mu$.

species. The Fock states for the decay process are given by

$$|i\rangle = |n_\pi(p), n_{\bar{l}}(k), n_h(q)\rangle \quad ; \quad |f\rangle = |n_\pi(p) - 1, n_{\bar{l}}(k) + 1, n_h(q) + 1\rangle \, . \tag{10.7.1}$$

In a similar fashion, the loss terms are obtained from the recombination $\bar{l}\nu_l \to \pi^+$ which has an initial state with $n_\pi(p), n_l(k), n_h(q)$ pion, charged lepton and neutrino quanta respectively. The final state is populated with $n_\pi(p) + 1, n_l(k) - 1, n_h(q) - 1$ of the appropriate quanta and the Fock states for the loss process are given by

$$|i\rangle = |n_\pi(p), n_{\bar{l}}(k), n_h(q)\rangle \quad ; \quad |f\rangle = |n_\pi(p) + 1, n_{\bar{l}}(k) - 1, n_h(q) - 1\rangle. \tag{10.7.2}$$

Where the relation between the neutrino mass and flavor eigenstates are given by (10.2.1):

A standard calculation yields the transition amplitudes at tree-level and the matrix elements for the gain term (decay) is given by ($\mathcal{H}_I$ is the interaction Hamiltonian density)

$$\begin{aligned} \mathcal{A}_{fi}|_{gain} &= -i \int d^4x \langle n_\pi(p) - 1, n_{\bar{l}}(k) + 1, n_h(q) + 1 | \mathcal{H}_I(x) | n_\pi(p), n_{\bar{l}}(k), n_h(q) \rangle \\ &= i\sqrt{2} G_F V_{ud} f_\pi H^*_{lh} \frac{2\pi}{\sqrt{V}} \frac{\bar{\mathcal{U}}_{\nu_h}(q,\sigma_1) \not{p} \mathcal{P}_L \mathcal{V}^l(k,\sigma_2)}{\sqrt{8 E_\pi(p) E_l(k) E_\nu(q)}} \\ &\times \delta_{\vec{p},\vec{k}+\vec{q}} \delta(E_\pi(p) - E_\nu(q) - E_l(k) \sqrt{n_\pi(p)} \sqrt{1 - n_l(k)} \sqrt{1 - n_\nu(q)} \end{aligned} \tag{10.7.3}$$

where $\mathcal{U}, \mathcal{V}$ are Dirac spinors and $\sigma_{1,2}$ are helicity quantum numbers, while the transition amplitude for the loss term (recombination) is given by

$$\begin{aligned}\mathcal{A}_{fi}|_{loss} &= -i\int d^4x\langle n_{\pi^+}(p)+1, n_{\bar{l}}(k)-1, n_h(q)-1|\mathcal{H}_I(x)|n_\pi(p), n_{\bar{l}}(k), n_h(q)\rangle \quad (10.7.4)\\ &= i\sqrt{2}\,G_F V_{ud} f_\pi H_{lh}\frac{2\pi}{\sqrt{V}}\frac{\bar{\mathcal{V}}^l(k,\sigma_2)\not{p}\mathcal{P}_L\mathcal{U}^{\nu_i}(q,\sigma_1)}{\sqrt{8E_\pi(p)E_l(k)E_\nu(q)}}\\ &\times\ \delta_{\vec{p},\vec{k}+\vec{q}}\delta(E_\pi(p)-E_\nu(q)-E_l(k))\sqrt{n_\pi(p)+1}\sqrt{n_l(k)}\sqrt{n_h(q)}\,.\end{aligned}$$

Summing over the final states leads to the transition probability per unit time (V is the quantization volume)

$$\begin{aligned}\frac{1}{\mathcal{T}}\sum_{\vec{k},\vec{p},\sigma_1,\sigma_2}\overline{|\mathcal{A}_{fi}|^2}_{gain} &= \frac{(2\pi)}{2E_h(q)}\int\frac{d^3p}{(2\pi)^3}\frac{2|H_{\alpha h}|^2|V_{ud}|^2G_F^2f_\pi^2}{2E_\pi(p)2E_l(k)}\,Tr[\not{p}\mathcal{P}_L(\not{q}+M_h)\not{p}\mathcal{P}_L(\not{k}-M_l)]\\ &\times\ n_\pi(p)(1-n_l(k))(1-n_\nu(q))\delta(E_\pi(p)-E_h(q)-E_l(k)) \quad (10.7.5)\end{aligned}$$

$$\begin{aligned}&= \frac{(2\pi)}{2E_h(q)}\int\frac{d^3p}{(2\pi)^3}\frac{2|H_{\alpha h}|^2|V_{ud}|^2G_F^2f_\pi^2}{2E_\pi(p)2E_l(k)}\,2[2(p\cdot q)(p\cdot k)-p^2(q\cdot k)]\\ &\times\ n_\pi(p)(1-n_l(k))(1-n_\nu(q))\delta(E_\pi(p)-E_\nu(q)-E_l(k)) \quad (10.7.6)\end{aligned}$$

where $\mathcal{T}$ is the total interaction time, not to be confused with temperature, and $k=|\vec{p}-\vec{q}|$. The loss term is calculated in the same way with the replacement $n_\pi\to 1+n_\pi$ and $1-n\to n$ for leptons. With these steps and explicitly evaluating the energy/momentum conservation leads to the quantum kinetic equation governing the sterile neutrino population:

$$\frac{dn_h(q;t)}{dt}=\frac{dn_h(q;t)}{dt}\Big|_{\rm gain}-\frac{dn_h(q;t)}{dt}\Big|_{\rm loss}\,, \quad (10.7.7)$$

where

$$\begin{aligned}\frac{dn_h(q;t)}{dt}\Big|_{\rm gain} &= \frac{2\pi}{2E_h(q)}\int\frac{d^3p}{(2\pi)^3}\,\frac{\overline{|\mathcal{M}_{fi}|^2}}{2E_\pi(p)2E_l(k)}\,n_\pi(p)(1-n_{\bar{l}}(k))(1-n_h(q;t))\\ &\times\ \delta(E_\pi(p)-E_l(k)-E_h(q))\ \ ;\ \ k=|\vec{p}-\vec{q}|\,, \quad (10.7.8)\end{aligned}$$

$$\begin{aligned}\frac{dn_h(q;t)}{dt}\Big|_{\rm loss} &= \frac{2\pi}{2E_h(q)}\int\frac{d^3p}{(2\pi)^3}\,\frac{\overline{|\mathcal{M}_{fi}|^2}}{2E_\pi(p)2E_l(k)}\,(1+n_\pi(p))n_{\bar{l}}(k)n_h(q;t)\\ &\times\ \delta(E_\pi(p)-E_l(k)-E_h(q))\ \ ;\ \ k=|\vec{p}-\vec{q}|\,, \quad (10.7.9)\end{aligned}$$

and the averaged $\overline{|\mathcal{M}_{fi}|^2}$ matrix element is given by

$$\overline{|\mathcal{M}_{fi}|^2} = 4|H_{\alpha h}|^2|V_{ud}|^2 G_F^2 f_\pi^2 \left[2(p\cdot q)(p\cdot k) - p^2(q\cdot k)\right]. \tag{10.7.10}$$

Therefore

$$\frac{dn_h(q;t)}{dt} = \Gamma^<(q)(1-n_h(q;t)) - \Gamma^>(q) n_h(q;t)\,, \tag{10.7.11}$$

with the gain and loss rates given by

$$\begin{aligned}
\Gamma^<(q) &= \frac{2\pi}{2E_h(q)} \int \frac{d^3p}{(2\pi)^3} \frac{\overline{|\mathcal{M}_{fi}|^2}}{2E_\pi(p) 2E_l(k)} n_\pi(p)(1-n_{\bar{l}}(k)) \\
&\times \delta(E_\pi(p) - E_l(k) - E_h(q)) \qquad (10.7.12) \\
\Gamma^>(q) &= \frac{2\pi}{2E_h(q)} \int \frac{d^3p}{(2\pi)^3} \frac{\overline{|\mathcal{M}_{fi}|^2}}{2E_\pi(p) 2E_l(k)} (1+n_\pi(p)) n_{\bar{l}}(k) \\
&\times \delta(E_\pi(p) - E_l(k) - E_h(q)) \;\; ; \;\; k = |\vec{p} - \vec{q}|\,. \qquad (10.7.13)
\end{aligned}$$

Performing the angular integration using the delta function constraint we find

$$\begin{aligned}
\Gamma^<(q) &= \frac{|H_{\alpha h}|^2|V_{ud}|^2 G_F^2 f_\pi^2}{8\pi} \frac{M_\pi^2(M_l^2+M_h^2) - (M_l^2 - M_\nu^2)^2}{qE_h(q)} \\
&\times \int_{p_-}^{p_+} \frac{dp\, p}{\sqrt{p^2+M_\pi^2}} \Big[n_\pi(p)(1-n_{\bar{l}}(|\vec{p}-\vec{q}|)) \Big] \qquad (10.7.14)
\end{aligned}$$

$$\begin{aligned}
\Gamma^>(q) &= \frac{|H_{\alpha h}|^2|V_{ud}|^2 G_F^2 f_\pi^2}{8\pi} \frac{M_\pi^2(M_l^2+M_h^2) - (M_l^2 - M_\nu^2)^2}{qE_h(q)} \\
&\times \int_{p_-}^{p_+} \frac{dp\, p}{\sqrt{p^2+M_\pi^2}} \Big[(1+n_\pi(p)) n_{\bar{l}}(|\vec{p}-\vec{q}|) \Big] \qquad (10.7.15)
\end{aligned}$$

The integration limits $p_\pm$ are obtained from the constraint

$$[(|\vec{p}| - |\vec{q}|)^2 + M_l^2]^{1/2} \leq E_\pi(p) - E_\nu(q) \leq [(|\vec{p}| + |\vec{q}|)^2 + M_l^2]^{1/2}. \tag{10.7.16}$$

The solutions are given by

$$p_\pm = \left| \frac{E_h(q)}{2M_h^2} [(M_\pi^2 - M_l^2 + M_h^2)^2 - 4M_\pi^2 M_h^2]^{1/2} \pm \frac{q(M_\pi^2 - M_l^2 + M_h^2)}{2M_h^2} \right|. \tag{10.7.17}$$

Note that these bounds coalesce at threshold when $M_\pi^2 - M_l^2 + M_h^2 = 2M_\pi^2 M_h^2$ and the population change vanishes as expected.

Using the relations

$$1 + n_\pi(p) = e^{E_\pi(p)/T}\, n_\pi(p) \quad ; \quad 1 - n_l(k) = e^{E_l(k)/T}\, n_l(k) \tag{10.7.18}$$

and using the energy delta function constraints, the detailed balance condition follows, namely

$$\Gamma^<(q) e^{E_h(q)/T} = \Gamma^>(q)\,. \tag{10.7.19}$$

These results are extended to cosmology by replacing the momentum with the physical momentum, $q \to Q_f = q_c/a(t)$ where physical energy and momentum is measured by an observer at rest with respect to the expanding spacetime.

10.7.2 Approximate Distributions.

The exact rate equation 10.4.24 is unwieldy and only limited analytical progress can be made. Recasting the rate equation in terms of the definitions of 10.4.20,10.4.26,10.4.27 and a dimensionless variable for the neutrino energy

$$\varepsilon \equiv E_h/T = \sqrt{y^2 + \frac{M_h^2}{M_\pi^2}\tau^2} = y\sqrt{1+u} \quad ; \quad u = m_h^2 \frac{\tau^2}{y^2} \tag{10.7.20}$$

leads to the following form of the rate equation which is more suitable to manipulations:

$$\begin{aligned} \frac{1}{\Lambda_{lh}}\frac{dn_{lh}}{d\varepsilon}(y,\varepsilon) &= \left(\frac{1}{m_h}\right)^3 \frac{1}{y}\frac{[(\varepsilon^2-y^2)^{1/2} - \frac{M_h^2}{6f_\pi^2}(\varepsilon^2-y^2)^{-1/2}]e^{-\varepsilon}}{1+e^{-\varepsilon}} \\ &\times \left\{ \ln\left(\frac{1-\exp\left(-\frac{1}{2m_h^2}(\delta_{lh}y + \Delta_{lh}\varepsilon)\right)}{1-\exp\left(-\frac{1}{2m_h^2}(-\delta_{lh}y + \Delta_{lh}\varepsilon)\right)}\right)\right. \\ &- \left. \ln\left(\frac{1+\exp\left(-\frac{1}{2m_h^2}(\delta_{lh}y + (\Delta_{lh}-2m_h^2)\varepsilon)\right)}{1+\exp\left(-\frac{1}{2m_h^2}(-\delta_{lh}y + (\Delta_{lh}-2m_h^2)\varepsilon)\right)}\right)\right\}. \end{aligned} \tag{10.7.21}$$

The population build up is obtained by integrating

$$n(\varepsilon_0, y) = \int_{\varepsilon_0}^{\infty} d\varepsilon' \frac{dn}{d\varepsilon}(\varepsilon', y) \quad ; \quad \varepsilon_0 = \sqrt{y^2 + m_h^2 \tau_0^2} = y\sqrt{1+u_0}\,. \tag{10.7.22}$$

where the initial population of ν_h has been neglected and the value of ε_0 is set by the temperature at which pions appear in thermal equilibrium, assumed almost immediately after the hadronization transition. It is assumed that pion thermalization happens instantaneously at the QCD phase transition and we set $\tau_0 \sim 1$ which is justified by the lattice results of [369] which suggest a continuous transition that allows for thermalization on strong interaction time scales.

The arguments of the exponentials inside of the logarithms can be shown to be < 0 so that the logarithms can be expanded consistently in a power series. Upon expanding the logarithms and integrating the rate, the population becomes

$$\frac{1}{\Lambda_{lh}} n_{lh}(\tau_0, y) = \left(\frac{1}{m_h}\right)^3 \frac{1}{y} \sum_{k=1}^{\infty} \frac{2}{k} \sinh\left(\frac{k\delta_{lh}}{2m_h^2} y\right) \left(\mathcal{I}_{lh}(k,y) - \frac{M_h^2}{6f_\pi^2} \mathcal{J}_{lh}(k,y)\right) \tag{10.7.23}$$

where the expressions $\mathcal{I}_{lh}(k), \mathcal{J}_{lh}(k)$ are given by

$$\begin{aligned} \mathcal{I}_{lh}(k,y) &= \frac{y^2}{2} \sum_{j=0}^{\infty} (-1)^j \left[\int_{u_0}^{\infty} du \frac{\sqrt{u}}{\sqrt{1+u}} \exp\left(-\left(j+1+\frac{k\Delta_{lh}}{2m_h^2}\right) y\sqrt{1+u}\right) \right. \\ &+ \left. (-1)^{k+1} \int_{u_0}^{\infty} du \frac{\sqrt{u}}{\sqrt{1+u}} \exp\left(-\left(j+1+\frac{k\Delta_{lh}}{2m_h^2} - k\right) y\sqrt{1+u}\right) \right] \end{aligned} \tag{10.7.24}$$

$$\begin{aligned} \mathcal{J}_{lh}(k,y) &= \frac{1}{2} \sum_{j=0}^{\infty} (-1)^j \left[\int_{u_0}^{\infty} du \frac{1}{\sqrt{u}\sqrt{1+u}} \exp\left(-\left(j+1+\frac{k\Delta_{lh}}{2m_h^2}\right) y\sqrt{1+u}\right) \right. \\ &+ \left. (-1)^{k+1} \int_{u_0}^{\infty} du \frac{1}{\sqrt{u}\sqrt{1+u}} \exp\left(-\left(j+1+\frac{k\Delta_{lh}}{2m_h^2} - k\right) y\sqrt{1+u}\right) \right] \\ &\quad . \end{aligned} \tag{10.7.25}$$

These expressions can be written in terms of the incomplete modified Bessel functions [432] but this proves to be an unilluminating exercise. However, it is possible that slightly simpler expressions can be found under appropriate approximations such as the assumption that the neutrinos are produced either ultra-relativistically or non-relativistically over the duration of production. The dimensionless variable $u = m_h^2 \tau^2/y^2$, which was introduced in Eq 10.7.20, is the obvious quantity which controls how much of an ultra-relativistic or non-relativistic dispersion relation is obeyed by the heavy neutrino; specifically $u \ll 1$ implies a light, ultra-relativistic species while $u \gg 1$ implies a very heavy, non-relativistic species.

From figs 58,59,60, it can be seen that y, τ take the values $1 < \tau < 10$ and $0 < y < 3$ from the beginning of production until decoupling which implies that $0 < y/\tau < 3$ for any neutrino produced. The production of purely ultra-relativistic neutrinos implies that $u \ll 1$ or that $m_\nu/m_\pi \ll y/\tau$ for all values of τ, y throughout production until freezeout. This assumption will hold for values of $m_\nu/m_\pi \ll 1$ or for very large values of y. For ultra-relativistic production, 10.7.24 10.7.25 simplify to

$$\begin{aligned} \mathcal{I}_{lh}(k,y) &- \frac{M_h^2}{6f_\pi^2}\mathcal{J}_{lh}(k,y)\Big|_{UR} = \left(\frac{1}{m_h}\right)^3 \frac{1+(-1)^{k+1}e^{ky}}{1+e^y} \exp\left(-\frac{k\Delta_{lh}}{2m_h^2}y\right) \\ &* \left(\frac{2\tau_0}{k\Delta_{lh}} \exp\left(-\frac{k\Delta_{lh}\tau_0^2}{4y}\right) + \left(\frac{1}{k\Delta_{lh}y}\right)^{1/2} \left(\frac{2y}{k\Delta_{lh}} - \frac{M_\pi^2}{6f_\pi^2}\right) \Gamma\left(1/2, \frac{k\Delta_{lh}\tau_0^2}{4y}\right)\right) . \end{aligned} \tag{10.7.26}$$

where $\Gamma(\nu, z)$ is the incomplete gamma function. Further assuming $M_h/M_\pi \ll 1$ leads to another simplification

$$2\sinh\left(\frac{k\delta_{lh}}{2m_h^2}y\right) \to \exp\left(\frac{k\Delta_{lh}}{2m_h^2}y\right) \exp\left(-ky/\Delta_{lh}\right) . \tag{10.7.27}$$

With these approximations, the form of the distribution for production of purely ultra-relativistic neutrinos becomes

$$n_{lh}(\tau_0, y)\Big|_{UR} = \frac{\Lambda_{lh}}{y^2} \sum_{k=1}^{\infty} \left[\frac{1+(-1)^{k+1}e^{ky}}{1+e^y}\right] \frac{\exp\left(-ky/\Delta_{lh}\right)}{k} \times J_k(\tau_0, y) \tag{10.7.28}$$

where

$$J_k(\tau_0, y) = 2\tau_0 \left(\frac{y}{k\Delta_{lh}}\right) \exp\left(-\frac{k\Delta_{lh}\tau_0^2}{4y}\right) + \left(\frac{y}{k\Delta_{lh}}\right)^{1/2} \left[\frac{2y}{k\Delta_{lh}} - \frac{M_\pi^2}{6f_\pi^2}\right] \Gamma\left(1/2, \frac{k\Delta_{lh}\tau_0^2}{4y}\right). \tag{10.7.29}$$

This expression is equivalent to the distribution which was obtained in [396] and explicitly depends on the time at which production begins, τ_0. In [396] it is shown that this approximate distribution is in excellent agreement with the distribution obtained from the exact numerical solution provided that $M_h \lesssim 1MeV$.

From figs 58,59,60, it is clear that for massive neutrinos with $M_h \gtrsim 10MeV$ there is a very differently shaped distribution function at small values of y compared with that seen from heavy neutrinos with masses below an MeV (as studied in [396]). This switch in shape between the distributions can be understood in terms of the production of *non-relativistic neutrinos* such that $u \gg 1$. The assumption that $u \gg 1$ implies that $m_\nu/m_\pi \gg y/\tau$ for all values of τ, y and, since $m_\nu/m_\pi < 1$, this approximation is clearly only valid for a range of values for $y \ll 1$ which explains the plateau in the distribution function for the heavier species. Under this assumption, the expressions 10.7.24,10.7.25 can be rewritten and are given below.

In the ultra-relativistic limit, exploiting $M_h \ll M_\pi$ was used in order to ignore several terms up to leading order in M_h/M_π; this allowed the sum, $\sum_j$, from 10.7.24,10.7.25 to be evaluated analytically. With heavier sterile neutrinos, such an approximation is unavailable and we are forced to retain the expression in the form of 10.7.30 if we desire any nontrivial momentum dependence. The non relativistic expressions are given here

$$\begin{aligned}
\mathcal{I}_{lh}(k,y)\Big|_{NR} &= \sum_{j=0}^{\infty}(-1)^j \frac{\exp\left[-(j+1+\frac{k\Delta_{lh}}{2m_h^2})y\sqrt{1+u_0}\right]}{(j+1+\frac{k\Delta_{lh}}{2m_h^2})^2}\Bigg[\\
&\left(1+(j+1+\frac{k\Delta_{lh}}{2m_h^2})y\sqrt{1+u_0}\right) + (-1)^{k+1}\, e^{ky\sqrt{1+u_0}} \\
&\times \left(\frac{j+1+\frac{k\Delta_{lh}}{2m_h^2}}{j+1+\frac{k\Delta_{lh}}{2m_h^2}-k}\right)^2 \left(1+(j+1+\frac{k\Delta_{lh}}{2m_h^2}-k)y\sqrt{1+u_0}\right)\Bigg] - \frac{y^2}{2}\mathcal{J}_k
\end{aligned}$$

$$
\begin{aligned}
\mathcal{J}_{lh}(k,y)\Big|_{NR} &= \sum_{j=0}^{\infty}\frac{(-1)^j}{2}\Bigg[\Big(\exp\left[(j+1+\frac{k\Delta_{lh}}{2m_h^2})y\right]E_1((j+1+\frac{k\Delta_{lh}}{2m_h^2})y(\sqrt{1+u_0}+1)) \\
&+ \exp\left[-(j+1+\frac{k\Delta_{lh}}{2m_h^2})y\right]E_1((j+1+\frac{k\Delta_{lh}}{2m_h^2})y(\sqrt{1+u_0}-1))\Big) \\
&+ (-1)^{k+1}\Big(\exp\left[(j+1+\frac{k\Delta_{lh}}{2m_h^2}-k)y\right]E_1((j+1+\frac{k\Delta_{lh}}{2m_h^2}-k)y(\sqrt{1+u_0}+1)) \\
&+ \exp\left[-(j+1+\frac{k\Delta_{lh}}{2m_h^2}-k)y\right]E_1((j+1+\frac{k\Delta_{lh}}{2m_h^2}-k)y(\sqrt{1+u_0}-1))\Big)\Bigg]
\end{aligned}
\tag{10.7.30}
$$

where $E_1(x)$ is the exponential integral.

11.0 PRODUCTION OF HEAVY STERILE NEUTRINOS FROM VECTOR BOSON DECAY AT $T \simeq M_W$.

Based on:

L. Lello, D. Boyanovsky, R. Pisarski, to be submitted

11.1 INTRODUCTION

The paradigm of standard cosmology premised on inflation plus cold dark matter, namely ΛCDM, is successful in describing large scale structure formation, however there seem to be discrepancies at smaller, galactic and sub-galactic scales. N-body simulations of cold dark matter produce dark matter profiles that generically feature cusps yet observations suggest a smooth-core profile [321, 322](core-cusp problem). The same type of N-body simulations also predict a large number of dark matter dominated satellites surrounding a typical galaxy which is inconsistent with current observations [323] (missing satellites problem). Both the missing satellites and core-cusp problem can be simultaneously resolved by allowing some fraction of the dark matter to be "warm" (WDM) [324, 325, 326, 327, 328, 329] with a massive "sterile" neutrino being a possible candidate [330, 69, 331, 433, 70, 332, 71, 320]. The "hotness" or "coldness" of a dark matter candidate is discussed in terms of its free streaming length, λ_{fs}, which is the cut-off scale in the linear power spectrum of density perturbations. Cold dark matter (CDM) features small ($\lesssim$ pc) λ_{fs} that brings about cuspy profiles whereas WDM features $\lambda_{fs} \sim$ few kpc which would lead to cored profiles. It

is yet unclear whether these discrepancies can still be explained within the standard cold dark matter model by including important baryonic physics in the cosmological simulations. Recent WDM simulations including velocity dispersion suggest the formation of cores but do not yet reliably constrain the mass of the WDM candidate in a model independent manner since the distribution function of these candidates is also an important quantity which determines the velocity dispersion and thereby the free streaming length[389].

In order to evade cosmological bounds [63] non-thermal distribution functions of WDM candidates are typically required. The mechanism of sterile neutrino production in the early universe through oscillations was originally studied in ref. [346] (see also the reviews [40]) and in [69, 347, 390, 391, 392] sterile neutrinos are argued to be a viable warm dark matter candidate produced out of local thermodynamic equilibrium (LTE) non-resonantly[346, 69, 390, 448] or resonantly[347] in presence of a lepton asymmetry. Models in which a scalar decays into a pair of sterile neutrinos at the electroweak scale (or higher) also yield an out-of-equilibrium distribution suitable for a sterile neutrino dark matter candidate [348, 349, 350, 356, 434]. Observations of the X-ray emission spectrum of the Andromeda galaxy with Chandra led to tight constraints on the model of sterile neutrinos produced non-resonantly[393], and more recently the report of observation of a 3.5 keV signal from the XMM Newton X-ray telescope has been argued to be due to a 7 keV sterile neutrino[351, 352], however this interpretation has been challenged [353, 354, 394, 395]. The prospect of a keV sterile DM candidate continues to motivate theoretical and observational studies [356, 355, 70, **?**, 357, 358, 349, 396, 397, 392, 398].

Neutrino masses, mixing and oscillations are uncontroversial evidence of physics beyond the standard model, and a robust experimental program has brought measurements of most of the parameters associated with light neutrino masses[334, 335] with several relevant questions poised to be answered in the near future [336]. Short baseline neutrino oscillation experiments (LSND, MiniBooNE) [43, 44] present a picture of the neutrino sector which may require an additional sterile neutrino species of mass $\sim 1eV$ [332, 343, 342] but there remains tension with other experiments [337] and a definitive resolution of these anomalies will require further experiments [340, 338, 139, 89, 399, 345]. An interpretation of short baseline experiment anomalies as a signal for sterile neutrinos leads to a relatively light mass

$\sim eV$ for the sterile neutrino which rules out this putative new sterile neutrino as a dark matter candidate. Howevever, many well motivated extensions beyond the standard model posit the existence of heavy neutrinos. It has been argued that sterile neutrinos with mass on the order of MeV or larger [124] can decay and could also explain the short baseline anomalies or, alternatively, heavy sterile neutrinos produced through rare decay channels could also explain the anomaly [400]. Recent proposals make the case for a program to search for heavy neutrinos [401, 402] in a wide range of experiments including hadron colliders [403, 404, 405, 406, 435]. Furthermore, it has been argued that heavy sterile neutrinos in the mass range $100 - 500\,\mathrm{MeV}$ can decay non thermally and enable evasion of cosmological and accelerator bounds[363]. Sterile neutrinos with mass $\simeq$ MeV can be of cosmological relevance in models of low reheating temperature[436]. A heavy sterile neutrino with mass $\simeq 14\,\mathrm{MeV}$ mixing angle $\theta \simeq 10^{-3}$ and lifetime $\tau_s \simeq 1.8 \times 10^5\,s$ has been proposed[437] as a novel solution to the "Lithium-problem", namely the nearly threefold discrepancy between the standard big-bang nucleosynthesis (BBN) prediction and observed abundance of ^7Li[438, 439, 440, 437, 441]. This solution relies on the energy injection provided by the decay of the sterile neutrino to destroy part of ^7Be prior to its conversion into ^7Li in the late stages of (BBN)[441, 437]. This mechanism has been recently re-analyzed and confirmed in ref.[442] with a sterile neutrino mass $\simeq 4.35\,\mathrm{MeV}$, mixing angle $\theta \leq 10^{-5} - 10^{-4}$ and lifetime $\simeq 1.8 \times 10^5\,s$. An important bonus of this mechanism, is that the decay of the $\simeq$ MeV (heavy) sterile neutrino yields an *increase* in the effective number of relativistic species $\Delta N_{eff} \simeq 0.34$ at the 95% CL[442]. The energy injection from the decay of heavy sterile neutrinos (with longer lifetime) may also contribute to early ionization[443]. Although there is no experimental evidence of heavy $M_s \geq 1\,\mathrm{MeV}$ sterile (like) neutrinos, there are stringent accelerator and cosmological bounds on their possible masses and mixing angles with active (like) neutrinos, these have been discussed in detail in refs.[98, 436, 444, 445].

The light active neutrinos feature a hierarchy of masses, separated by nearly two orders of magnitude between solar and atmospheric. Possible extensions beyond the standard model may also accommodate a hierarchy of *heavy* neutrinos[320, 390, 446]. Current and future underground neutrino detectors may be able to probe dark matter candidates with $\simeq$ few MeV[410]. The possibility of a hierarchy of *heavy* sterile neutrinos offers novel pro-

duction mechanisms of warm (and hot) dark matter via the cascade decays of the heavier mass eigenstates to the lower mass states in the hierarchy. This possibility would be akin to models of many dark matter components proposed recently[411], wherein the decay of one heavy component can seed the production of a lighter component. This possibility would lead to the scenario of *mixed dark matter* described by several species of massive neutrinos with non-equilibrium distribution functions, such scenarios would evade Lyman-α constraints[412]. A recent article[447] argued on various possible production mechanisms of sterile neutrino directly from standard model processes available throughout the thermal history of the universe and analyzed in detail the scenario of production of mixed dark matter (colder, warmer and hotter) from pion decay shortly after the QCD crossover. This analysis along with previous work[396] also suggested (but did not quantify) that the decay of a heavier sterile species into the lighter active neutrinos would yield an increase in N_{eff}, a suggestion that seems to have been validated by the most recent analysis in ref.[442] within the context of energy injection post (BBN).

Motivation and Goals: Sterile neutrinos with masses in the range KeV − few MeV may play a very important role in cosmology. Most of the studies of their production and freeze-out have focused on the well motivated mass range of few KeV as possible warm dark matter candidates. However, if the hierarchy of masses and mixing of light active neutrinos is of any guide in extensions beyond the standard model, a possible hierarchy of heavier sterile neutrinos that mix with the light active neutrinos with very small mixing angles may emerge. In this scenario, the possibility that heavier neutrinos yield a mixture of dark matter components, from cold (heavier) to warmer (lighter) or that the decay of heaviest sterile (like) neutrinos lead to lighter dark matter species with important cosmological impact, motivates our study of the production and freeze out of sterile neutrinos in a wider range of masses and temperatures. In this article we study the production of sterile (like) neutrinos solely from *standard model interactions*, ref.[447] identified several possible processes available throughout the thermal history of the Universe that *may* lead to the production of a sterile species from its mixing with active neutrinos.

Our goals in this article are twofold:

- **i:)** We seek to provide a consistent quantum kinetics description of production and freeze-out valid in a wide range of temperature and general expressions for production rates and effective mixing angle in the medium under a minimal set of assumptions. These are: **a)** sterile neutrinos couple to active ones via a see-saw type mass matrix, and we consider only standard model interaction vertices. **b)** consistently with accelerator and cosmological bounds[98, 436, 393, 351, 352, 357, 444, 445] we consider the vacuum mixing angle $\theta \ll 1$. Taken together, these bounds suggest that $\theta^2 \lesssim 10^{-5}$ for a wide range of masses $M_s \lesssim 300\,\mathrm{MeV}$. **c):** The validity of the perturbative expansion in standard model couplings and that standard model degrees of freedom (including active-like neutrinos) are in local thermodynamic equilibrium (LTE) during the relevant time scales for production and freeze-out. This latter assumption is consistent with (actually a corollary of) $\theta \ll 1$.
- **ii:)** We study in detail the production of sterile neutrinos to leading order in standard model coupling α_w and without restriction to a particular mass scale of the sterile-like degrees of freedom. To leading order $\mathcal{O}(\alpha_w)$ production of the sterile-like mass eigenstates occurs from the decay of the massive vector bosons W, Z in the medium. We focus on the temperature scale $T \simeq M_w$ sufficiently below the electroweak crossover at $T_{ew} \simeq 160\mathrm{GeV}$[391] so that the Higgs mechanism is fully operative and the masses of the W, Z are nearly the zero temperature values. In this temperature regime, W, Z are in local thermodynamic equilibrium (LTE) in the medium and the possible finite temperature corrections to their masses and couplings are subleading and can be neglected, particularly for the production of sterile -like neutrinos which is further suppressed by the effective mixing angle in the medium.

Brief summary of results: We consider a model of one active and one sterile neutrino to extract the main aspects of production and freeze out in simple scenario. The main results are the following:

- We obtain the mass eigenstates, effective mixing angles and damping rates directly from the equations of motion in the medium in terms of the full self-energy *to all orders in weak interactions*. We give an expression for the effective mixing angles which is broadly

valid for $\theta \ll 1$ and to all orders in perturbation theory in standard model couplings and at any temperature. The mixing angle in the medium depends strongly on helicity: negative helicity neutrinos (and positive helicity antineutrinos) feature mixing angles that are strongly suppressed at high temperature, whereas for positive helicity neutrinos (and negative helicity anti-neutrinos) the corrections to the mixing angle are subleading and the effective mixing angle is nearly the same as the vacuum mixing angle. This is a consequence of the fact that the interaction of positive helicity neutrino with the medium is helicity suppressed. Damping rates are also strongly dependent on helicity and suppressed for positive helicity neutrinos. We obtain the general form of the quantum kinetic equation that describes production and freeze out of sterile-like neutrinos to *all orders in standard model coupling* and to leading order in $\theta \ll 1$. The production rate is completely determined by the damping rate of sterile-like *mass eigenstates* and the mixing angle in the medium. Although the production rate of positive helicity states features a helicity suppression, it is comparable to the rate for the negative helicity states in a wide regime of masses because of the large suppression of the mixing angle in the medium for negative helicity states which are more strongly coupled to the plasma.

- For sterile-like masses $M_s \ll M_w$ we find a Mikheyev-Smirnov-Wolfenstein (MSW)[30] resonance in *absence of a leptonic asymmetry*, which however, is screened by the damping rate and does *not* lead to enhanced production. For this mass range of M_s negative helicity states freeze out at $T_f^- \simeq 5\,\mathrm{GeV}$ whereas positive helicity states freeze-out at $T_f^+ \simeq 8\mathrm{GeV}$. Both feature highly non-thermal distribution functions with the broader and hotter distribution for the negative helicity states and a sharper and colder distribution for the positive helicity states. Paradoxically, this is a consequence of a *longer freeze-out time* for the negative helicity states despite the fact that their coupling to the environment is stronger. This is a surprising result stemming from a competition between a diminishing damping rate and an *increasing* effective mixing angle as the temperature diminishes. We argue that this leading order production mechanism establishes a *lower bound* for the abundance. We find however, that sterile-like neutrinos produced via vector boson decay do not satisfy the various bounds on lifetimes and mixing angles to be viable keV dark matter candidates. However they can be suitable as the MeV

sterile neutrinos that are conjectured to provide a solution to the ^7Li problem, with the caveat that there seems to be tension among the various bounds available in the literature[442, 445]. Just as these heavier neutrinos may decay injecting energy into the plasma as the solution to this problem, we also conjecture that they may also decay into lighter $\simeq$ KeV sterile neutrinos (with a much smaller branching ratio) that could be suitable WDM candidates.

To the best of our knowledge there has not yet been a systematic study of sterile (like) neutrino production at the scale $T \simeq 100$GeV with cosmological expansion. Our study is motivated by the possible cosmological relevance of sterile neutrinos in a wide range of masses, and complements previous studies that focus on lower temperature regimes.

11.2 MASS EIGENSTATES, DAMPING RATES AND MIXING ANGLES IN THE MEDIUM:

We consider the Standard Model with only one leptonic generation: one active neutrino and its charged lepton partner and one sterile neutrino to discuss the main aspects in the simplest setting. The active and sterile neutrinos only interact via a see-saw type mass matrix, the Lagrangian density is

$$\mathcal{L} = \mathcal{L}_{SM} + \overline{\nu}_s \, i \partial\!\!\!/ \nu_s - \overline{\nu}_\alpha \; \mathbb{M}_{\alpha\beta} \; \nu_\beta + \text{h.c} \; ; \; \alpha, \beta = a, s \; , \tag{11.2.1}$$

where a, s refer to active and sterile respectively and

$$\mathbb{M} = \begin{pmatrix} 0 & m \\ m & M_s \end{pmatrix} . \tag{11.2.2}$$

We will neithre specify nor model the origin of this mass matrix simply assuming the standard see-saw form. Introducing the "flavor" doublet (ν_a, ν_s) the diagonalization of the mass term

$\mathbb{M}$ is achieved by a unitary transformation to the mass basis (ν_1, ν_2), namely

$$\begin{pmatrix} \nu_a \\ \nu_s \end{pmatrix} = U(\theta) \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} \;\; ; \;\; U(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} , \tag{11.2.3}$$

where

$$\cos(2\theta) = \frac{M_s}{\left[M_s^2 + 4m^2\right]^{\frac{1}{2}}} \;\; ; \;\; \sin(2\theta) = \frac{2m}{\left[M_s^2 + 4m^2\right]^{\frac{1}{2}}} \; . \tag{11.2.4}$$

In the mass basis

$$\begin{aligned} \mathbb{M}_{diag} &= U^{-1}(\theta)\, \mathbb{M}\, U(\theta) = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \\ M_1 &= \frac{1}{2}\left[M_s - \left[M_s^2 + 4m^2\right]^{\frac{1}{2}}\right] \;\; ; \;\; M_2 = \frac{1}{2}\left[M_s + \left[M_s^2 + 4m^2\right]^{\frac{1}{2}}\right] \; . \end{aligned} \tag{11.2.5}$$

We focus on the case $m \ll M_s$, therefore

$$M_1 \simeq -\frac{m^2}{M_s} \;\; ; \;\; M_2 \simeq M_s \;\; ; \;\; \sin(2\theta) \simeq 2\theta \simeq \frac{2m}{M_s} \ll 1 \; . \tag{11.2.6}$$

We work in unitary gauge which exhibits the physical degrees of freedom of massive vector bosons in thermodynamic equilibrium. The equations of motion were derived previously in references[416, 449, 115]. In particular, ref.[416] also includes contributions from Yukawa couplings between the sterile neutrino and scalar fields, but we will *not* consider such extension here as this implies a particular model for the origin of the mass matrix. Our focus here is to study the sterile neutrino production solely from standard model interactions (charged and neutral currents) under the minimal set of assumptions discussed above.

Introducing the flavor doublet $\Psi^T = (\nu_a, \nu_s)$ the equation of motion *in the flavor basis* is[416, 449, 115](for details see the appendix in ref.[449]).

$$\left(i \not{\partial} \mathbb{1} - \mathbb{M} + \mathbf{\Sigma}^t \mathbb{L}\right) \Psi(\vec{x}, t) + \int d^3x' dt' \;\; \mathbf{\Sigma}^r(\vec{x} - \vec{x}', t - t')\, \mathbb{L}\, \Psi(\vec{x}', t') = 0 \tag{11.2.7}$$

where $\mathbb{1}$ is the identity matrix in flavor space and $\mathbb{L} = (1 - \gamma^5)/2$ is the left-handed chiral projection operator. The full one-particle irreducible self-energy includes local tadpole (Σ^t)

and non-local dispersive ($\Sigma^r(\vec{x}-\vec{x}',t-t')$) contributions. It is solely arising from *standard model interactions to all orders* and is diagonal in the flavor basis, namely

$$\mathbf{\Sigma} \equiv \Sigma \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{11.2.8}$$

Furthermore in factoring out the projector $\mathbb{L}$, the remaining self-energy is calculated in the vector-like theory. For example, the one loop contributions to the self-energy are shown in fig.(70), this is the leading order contribution to the self energy.

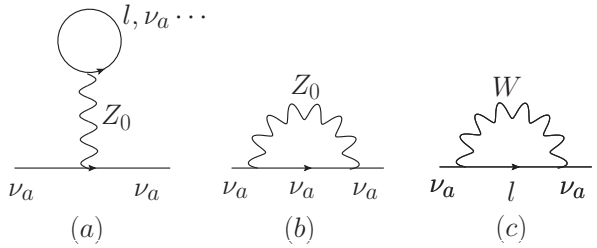


Figure 70: One loop contributions to the self-energy. The neutral current tadpole is proportional to the lepton (and quark) asymmetry.

Introducing the space-time Fourier transform in a spatial volume V

$$\Psi(\vec{x},t) = \frac{1}{\sqrt{V}} \sum_{\vec{q}} \int d\omega \, e^{i\vec{q}\cdot\vec{x}} \, e^{-i\omega t} \tilde{\Psi}(\omega,\vec{q}) \tag{11.2.9}$$

and similarly for the self-energy kernels, the effective Dirac equation in the flavor basis becomes[416, 449, 115] (see the appendix in **??**)

$$\left[\left(\gamma_0\omega - \vec{\gamma}\cdot\vec{q}\right)\mathbb{1} - \mathbb{M} + \left(\mathbf{\Sigma}^t + \mathbf{\Sigma}(\omega,\vec{q})\right)\mathbb{L} \right] \tilde{\Psi}(\omega,\vec{q}) = 0\,, \tag{11.2.10}$$

this equation of motion is *exact*, since the self-energy includes all order contributions in standard model couplings. The bracket in (11.2.10) is the inverse or the retarded propagator whose poles in the complex plane determine the dispersion relations and damping rates of *the mass eigenstates in the medium.*

The tadpole contribution $\mathbf{\Sigma}^t$ is local therefore it is independent of $\omega, \vec{q}$ and in (11.2.10) $\Sigma(\omega, \vec{q})$, is the space-time Fourier transform of $\Sigma^r(\vec{x} - \vec{x}', t - t')$ and features a dispersive representation[416, 449, 115]

$$\Sigma(\omega, \vec{q}) = \frac{1}{\pi} \int_{-\infty}^{\infty} dq_0 \, \frac{\mathrm{Im}\Sigma(q_0, \vec{q})}{q_0 - \omega - i\, 0^+} \, . \tag{11.2.11}$$

From this dispersive form it follows that

$$\Sigma(\omega, \vec{q}) = \mathrm{Re}\Sigma(\omega, \vec{q}) + i\, \mathrm{Im}\Sigma(\omega, \vec{q}) \tag{11.2.12}$$

with

$$\mathrm{Re}\Sigma(\omega, \vec{q}) = \frac{1}{\pi} \int_{-\infty}^{\infty} dq_0 \, \mathcal{P} \left[\frac{\mathrm{Im}\Sigma(q_0, \vec{q})}{q_0 - \omega} \right] . \tag{11.2.13}$$

The real part yields the "index of refraction" in the medium and the imaginary part determines the *damping rate* of the single (quasi) particle excitations. The tadpole term must be calculated separately and does not feature a dispersive representation.

Although in this article we will focus on the one-loop contributions to the self-energy from standard model charged and neutral current interactions, the form of the equations of motion and the dispersive form of the self-energy (not the tadpole) is generally valid in principle to *all orders* in standard model interactions which are of the $V - A$ form.

A subtle but important conceptual issue arises in the neutral current contribution to the self-energy with internal loop propagators for neutrinos. The propagators correspond to mass eigenstates, therefore in principle the perturbative loop expansion should be carried out in the *mass basis* rather than in the flavor basis. Furthermore, if the neutrino propagators describe neutrinos thermalized in the medium in terms of the equilibrium Fermi-Dirac distribution function, not only these propagators correspond to mass (energy) eigenstates but also are *assumed* to be in thermal equilibrium. We will assume the following: i) very small mixing angle $\theta \ll 1$ so that to leading order in this mixing angle, the active-like mass eigenstate can be taken to be the active flavor eigenstate, ii) in the temperature regime of interest in this article $T \simeq M_{w,z}$ active (flavor) neutrinos are in (local) thermal equilibrium. Under these assumptions (the validity of which will be confirmed later) we consider the internal loop

propagators in the neutral current contribution to be those of active neutrinos in thermal equilibrium to leading order in the mixing angle.

As a consequence of the $V-A$ nature of the standard model couplings of neutrinos, $\Sigma(\omega,\vec{q})$ has the general form of a vector-like theory

$$\Sigma^t + \Sigma(\omega,\vec{q}) \equiv \gamma^0\, A(\omega,\vec{q}) - \vec{\gamma}\cdot\widehat{\mathbf{q}}\; B(\omega,\vec{q})\,, \tag{11.2.14}$$

and

$$\mathbf{\Sigma}^t + \mathbf{\Sigma}(\omega,\vec{q}) = \gamma^0\, \mathbb{A}(\omega,\vec{q}) - \vec{\gamma}\cdot\widehat{\mathbf{q}}\; \mathbb{B}(\omega,\vec{q}) \tag{11.2.15}$$

where in the flavor basis

$$\mathbb{A}(\omega,\vec{q}) = \begin{pmatrix} A(\omega,\vec{q}) & 0 \\ 0 & 0 \end{pmatrix} \;\;;\;\; \mathbb{B}(\omega,\vec{q}) = \begin{pmatrix} B(\omega,\vec{q}) & 0 \\ 0 & 0 \end{pmatrix}, \tag{11.2.16}$$

The equations of motion simplify by projecting with $\mathbb{L}=(1-\gamma^5)/2; \mathbb{R}=(1+\gamma^5)/2$ and expanding in helicity eigenstates. Following the steps of ref.[416] we write for the left (L) and right (R) fields

$$\Psi_L = \sum_{h=\pm1} v^h \otimes \Psi_L^h \;\;;\;\; \Psi_L^h = \begin{pmatrix} \nu_a^h \\ \nu_s^h \end{pmatrix}_L, \tag{11.2.17}$$

and

$$\Psi_R = \sum_{h=\pm1} v^h \otimes \Psi_R^h \;\;;\;\; \Psi_R^h = \begin{pmatrix} \nu_a^h \\ \nu_s^h \end{pmatrix}_R, \tag{11.2.18}$$

where the left and right handed doublets are written in the *flavor* basis, and v^h are eigenstates of the helicity operator

$$\widehat{h}(\widehat{\mathbf{k}}) = \gamma^0\vec{\gamma}\cdot\widehat{\mathbf{k}}\,\gamma^5 = \vec{\sigma}\cdot\widehat{\mathbf{k}}\begin{pmatrix} \mathbb{1} & 0 \\ 0 & \mathbb{1} \end{pmatrix} \tag{11.2.19}$$

namely,

$$\vec{\sigma}\cdot\widehat{\mathbf{k}}\; v^h = h\; v^h \;\;;\;\; h=\pm1\,. \tag{11.2.20}$$

We find in the *flavor* basis the equation of motion for the left and right-handed component doublets

$$\Big[(\omega^2-q^2)\mathbb{1} + \big(\omega - hq\big)\big(\mathbb{A}+h\mathbb{B}\big) - \mathbb{M}^2\Big]\Psi_L^h = 0 \tag{11.2.21}$$

and

$$\Big[\omega - h\,q\Big]\Psi_R^h = \mathbb{M}\,\gamma^0\,\Psi_L^h\,, \tag{11.2.22}$$

where $\mathbb{M}$ is given by (11.2.2) and

$$\mathbb{M}^2 = \overline{M}^2\mathbb{1} + \frac{\delta M^2}{2}\begin{pmatrix} -\cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & \cos(2\theta) \end{pmatrix}. \tag{11.2.23}$$

with

$$\overline{M}^2 \equiv \frac{1}{2}\big(M_1^2 + M_2^2\big) \;\;;\;\; \delta M^2 \equiv M_2^2 - M_1^2\,, \tag{11.2.24}$$

and $M_{1,2}$ are given by eqn. (11.2.6). The results (11.2.21,11.2.22) are *general* for standard model couplings of the active (flavor) neutrinos and sterile neutrinos that only interact with active ones via a see-saw type mass matrix. Before discussing in detail the one-loop contribution from charged and neutral currents, we want to establish general results for the effective mixing angle in the medium and damping rates. The operator on the left hand side of (11.2.21)

$$\Big[\mathbb{S}_L^h(\omega,q)\Big]^{-1} = \Big[(\omega^2-q^2)\mathbb{1} + \big(\omega - hq\big)\big(\mathbb{A} + h\mathbb{B}\big) - \mathbb{M}^2\Big]\,, \tag{11.2.25}$$

defines the inverse propagator in the flavor basis for the left handed component projected on helicity eigenstates. The correct "mass eigenstates" correspond to the (complex) poles of $\mathbb{S}$, the real part describes the correct propagating frequencies and the imaginary parts describe the *damping rate* of single (quasi) particle excitations. We will extract these "mass eigenstates" from the (complex) zeroes of $[\mathbb{S}_L^h]^{-1}$ by invoking the following approximations whose validity will be assessed below:

- Ultrarelativistic approximation: $q \gg M_s$, this entails that the produced sterile like neutrinos freeze out while relativistic.
- $\theta \ll 1$, in particular we will assume that the self-energy correction is larger in magnitude than the vacuum mixing angle. The precise condition will be discussed below.
- Validity of the perturbative expansion, in particular that the self-energy corrections are smaller than the unperturbed dispersion relations. This assumption will be clarified and discussed in detail in the analysis that follows.

Introducing

$$\Omega^h \equiv (\omega - h\,q)\left(A(\omega,q) + h\,B(\omega,q)\right), \tag{11.2.26}$$

and using (11.2.23) we obtain

$$\begin{aligned}\left[\mathbb{S}^h_L(\omega,q)\right]^{-1} &= \left(\omega^2 - q^2 - \overline{M}^2 + \frac{\Omega^h}{2}\right)\mathbb{1} \\ &\quad -\frac{1}{2}\quad \sqrt{\left(\delta M^2\cos(2\theta)+\Omega^h\right)^2 + \left(\delta M^2\sin(2\theta)\right)^2}\ \begin{pmatrix} -\mathbb{C}^h_m(\omega,q) & \mathbb{S}^h_m(\omega,q) \\ \mathbb{S}^h_m(\omega,q) & \mathbb{C}^h_m(\omega,q)\end{pmatrix}\end{aligned} \tag{11.2.27}$$

where

$$\mathbb{C}^h_m(\omega,q) = \frac{\delta M^2\cos(2\theta)+\Omega^h(\omega,q)}{\sqrt{\left(\Omega^h(\omega,q)+\delta M^2\cos(2\theta)\right)^2 + \left(\delta M^2\ \sin(2\theta)\right)^2}} \tag{11.2.28}$$

$$\mathbb{S}^h_m(\omega,q) = \frac{\delta M^2\sin(2\theta)}{\sqrt{\left(\Omega^h(\omega,q)+\delta M^2\cos(2\theta)\right)^2 + \left(\delta M^2\ \sin(2\theta)\right)^2}}, \tag{11.2.29}$$

with

$$(\mathbb{C}^h_m(\omega,q))^2 + (\mathbb{S}^h_m(\omega,q))^2 = 1\,. \tag{11.2.30}$$

If the imaginary part of the self energy vanishes (or can be neglected) then

$$\mathbb{C}^h_m(\omega,q) = \cos(2\theta_m)\quad ;\quad \mathbb{S}^h_m(\omega,q) = \sin(2\theta_m)\,, \tag{11.2.31}$$

and θ_m would be the mixing angle in the medium. However, the absorptive (imaginary) part of the self-energy (related to the damping rate of quasi particle excitations) prevent such identification.

The matrix

$$\begin{pmatrix} -\mathbb{C}^h_m(\omega,q) & \mathbb{S}^h_m(\omega,q) \\ \mathbb{S}^h_m(\omega,q) & \mathbb{C}^h_m(\omega,q)\end{pmatrix}$$

has null trace and determinant (-1) as a consequence of (11.2.30), therefore *real* eigenvalues $\lambda = \pm 1$ with the following eigenvectors:

$$\begin{pmatrix} c^h(\omega,q) \\ -s^h(\omega,q) \end{pmatrix} \quad ; \quad \lambda = -1 \tag{11.2.32}$$

$$\begin{pmatrix} s^h(\omega,q) \\ c^h(\omega,q) \end{pmatrix} \quad ; \quad \lambda = 1 \tag{11.2.33}$$

where

$$c^h(\omega,q) \;=\; \left[\frac{1+\mathbb{C}^h_m(\omega,q)}{2}\right]^{1/2} \tag{11.2.34}$$

$$s^h(\omega,q) \;=\; \left[\frac{1-\mathbb{C}^h_m(\omega,q)}{2}\right]^{1/2} . \tag{11.2.35}$$

For vanishing absorptive part $s^h \equiv \sin(\theta_m), c^h \equiv \cos(\theta_m)$ with θ_m the (real) mixing angle in the medium.

To leading order for $\theta \ll 1$ and $M_1 \ll M_2 \simeq M_s$ we obtain the following eigenvalues of $\Big[\mathbb{S}\Big]^{-1}$

$$S^{-1} \;\simeq\; \omega^2 - q^2 - M_2^2 - \frac{\theta^2\,(M_2^2)^2}{\Big(M_2^2 + \Omega^h(\omega,q)\Big)} + \theta^2 M_2^2$$

$$\text{for } \lambda = +1 \quad ; \quad \text{eigenvector} \begin{pmatrix} s^h(\omega,q) \\ c^h(\omega,q) \end{pmatrix} \tag{11.2.36}$$

$$S^{-1} \;\simeq\; \omega^2 - q^2 - M_1^2 + \Omega^h(\omega,q) + \frac{\theta^2\,(M_2^2)^2}{\Big(M_2^2 + \Omega^h(\omega,q)\Big)} - \theta^2 M_2^2$$

$$\text{for } \lambda = -1 \quad ; \quad \text{eigenvector} \begin{pmatrix} c^h(\omega,q) \\ -s^h(\omega,q) \end{pmatrix} . \tag{11.2.37}$$

It is clear that the eigenvector corresponding to eigenvalue $+1$ corresponds to a sterile-like neutrino in the medium: the radiative correction (self-energy) enters solely with the mixing angle and vanishes for vanishing mixing angle, whereas the eigenvector corresponding to eigenvalue -1 is active-like, with radiative correction (Ω^h) even for $\theta = 0$. The inverse

of $\left[\mathbb{S}_L^h(\omega,q)\right]^{-1}$ is the inverse of the propagator, therefore its complex zeroes describe the complex poles. Under the validity of perturbation theory (discussed below in detail) we write

$$\omega = \omega_j(q) + \delta\omega_j^h \;\;;\;\; \omega_j(q) = \sqrt{q^2 + M_j^2} \simeq q + \frac{M_j^2}{2q} \;\;;\;\; j = 1,2 \tag{11.2.38}$$

in the relativistic approximation and introducing (in the relativistic limit)

$$\Delta_j^h(q) + i\gamma_j^h(q) = \frac{\Omega^h(\omega_j,q)}{2q} \simeq \frac{(\omega_j - h\,q)}{2q}\left[A(\omega = q,q) + h\,B(\omega = q,q)\right], \tag{11.2.39}$$

with $j = 2$ for sterile-like (eigenvalue $\lambda = 1$) and $j = 1$ for active-like (eigenvalue $\lambda = -1$) where both Δ and γ are real, and introducing

$$\xi = \frac{M_s^2}{2q} \tag{11.2.40}$$

we find the position of the poles in the propagator ("mass eigenstates") at

$$\begin{aligned} \delta\omega_2^h &= \frac{\theta^2\left(\xi + \Delta_2^h(q) - i\gamma_2^h(q)\right)}{\left[\left(1 + \frac{\Delta_2^h(q)}{\xi}\right)^2 + \left(\frac{\gamma_2^h(q)}{\xi}\right)^2\right]} - \theta^2\xi \\ &\text{for } \lambda = +1 \;\;;\;\; \text{eigenvector} \begin{pmatrix} s^h(\omega = q,q) \\ c^h(\omega = q,q) \end{pmatrix} \end{aligned} \tag{11.2.41}$$

for the "sterile-like" neutrino and

$$\begin{aligned} \delta\omega_1^h &= -\left(\Delta_1^h(q) + i\gamma_1^h(q)\right) - \frac{\theta^2\left(\xi + \Delta_1^h(q) - i\gamma_1^h(q)\right)}{\left[\left(1 + \frac{\Delta_1^h(q)}{\xi}\right)^2 + \left(\frac{\gamma_1^h(q)}{\xi}\right)^2\right]} + \theta^2\xi \\ &\text{for } \lambda = -1 \;\;;\;\; \text{eigenvector} \begin{pmatrix} c^h(\omega = q,q) \\ -s^h(\omega = q,q) \end{pmatrix} \end{aligned} \tag{11.2.42}$$

for the "active-like" neutrino.

We now introduce the *effective mixing angle in the medium*

$$\theta^h_{eff}(q) = \frac{\theta}{\left[\left(1+\frac{\Delta^h_j(q)}{\xi}\right)^2 + \left(\frac{\gamma^h_j(q)}{\xi}\right)^2\right]^{1/2}}\,, \tag{11.2.43}$$

for each mass eigenstate $j = 1, 2$, in terms of which the position of the (quasi) particle poles (11.2.42,11.2.43) are written as

$$\delta\omega^h_2 = \Delta^h_2(q)\,\left(\theta^h_{eff}(q)\right)^2 + \xi\left[\left(\theta^h_{eff}(q)\right)^2 - \theta^2\right] - i\,\gamma^h_2(q)\left(\theta^h_{eff}(q)\right)^2 \;\; (\text{sterile}-\text{like})\,, \tag{11.2.44}$$

$$\delta\omega^h_1 = -\Delta^h_1(q)\,\left[1+\left(\theta^h_{eff}(q)\right)^2\right] - \xi\left[\left(\theta^h_{eff}(q)\right)^2 - \theta^2\right] - i\,\gamma^h_1(q)\left[1-\left(\theta^h_{eff}(q)\right)^2\right] \;\; (\text{active}-\text{like})\,. \tag{11.2.45}$$

Writing

$$\delta\omega^h_j = \Delta E^h_j - i\frac{\Gamma^h_j}{2}\,, \tag{11.2.46}$$

for the corresponding helicity component, the imaginary part Γ yields the damping rate for the single (quasi) particle excitations in the medium, namely the "mass eigenstates" in the medium evolve in time as

$$\nu^h_j(q) \simeq e^{-i\omega_j t}\, e^{-i\Delta E^h_j t}\, e^{-\Gamma^h_j t/2} \Rightarrow |\nu^h_j(q)|^2 \simeq e^{-\Gamma^h_j t} \;\; ; \;\; j = 1, 2\,, \tag{11.2.47}$$

where the damping rates Γ^h_j are given by

$$\Gamma^h_2 \;=\; 2\gamma^h(q)\left(\theta^h_{eff}(q)\right)^2 \simeq 2\gamma^h(q)\,\sin^2(\theta^h_{eff}) \;\; \text{sterile}-\text{like} \tag{11.2.48}$$

$$\Gamma^h_1 \;=\; 2\gamma^h(q)\left[1-\left(\theta^h_{eff}(q)\right)^2\right] \simeq 2\gamma^h(q)\,\cos^2(\theta^h_{eff}) \;\; \text{active}-\text{like}\,. \tag{11.2.49}$$

Even when a particle cannot decay in the vacuum, the spectral density may feature a width in the medium as a consequence of dissipative processes arising from the coupling to excitations in the medium. In this case the width describes the relaxation of the quasi-particle in linear response[450, 451, 421].

The coefficient $\omega - hq$ in (11.2.26) is noteworthy: for positive ω the positive helicity component $h = 1$ is *helicity* suppressed, on the mass shell of the (vacuum) mass eigenstates in the relativistic limit $\omega - q \simeq M^2_{1,2}/2q$. This is the usual helicity suppression from the $V-A$

form of the interaction and has important consequences: the damping rate for *positive helicity* sterile-like neutrinos is much smaller than that for the negative helicity, and the medium corrections to the mixing angle are also *suppressed* for the positive helicity component. This suppression will have important and unexpected consequences on the rate of production of the sterile-like species as will be discussed below in detail.

The functions $\gamma^{\pm}(q)$ require the combinations $A \mp B$ that define the self energy (11.2.14), these combinations are handily extracted as follows: introduce the four-vectors

$$Q^{-}_{\mu} = \frac{1}{q}(q, -\vec{q}) \quad ; \quad Q^{+}_{\mu} = \frac{1}{q}(q, \vec{q}) \tag{11.2.50}$$

and defining

$$\Sigma^{-}(q) \equiv A(\omega = q, q) - B(\omega = q, q) \quad ; \quad \Sigma^{+}(q) \equiv A(\omega = q, q) + B(\omega = q, q) \tag{11.2.51}$$

we obtain

$$\Big(A(\omega, q) - B(\omega, q)\Big)_{\omega=q} = \frac{1}{4}\,\mathrm{Tr}\not{Q}^{-}\,\Sigma(q, q) \equiv \Sigma^{-}(q)\,, \tag{11.2.52}$$

$$\Big(A(\omega, q) + B(\omega, q)\Big)_{\omega=q} = \frac{1}{4}\,\mathrm{Tr}\not{Q}^{+}\,\Sigma(q, q) \equiv \Sigma^{+}(q)\,, \tag{11.2.53}$$

$$\gamma^{-}(q) = \mathrm{Im}\,\Sigma^{-}(q) \quad ; \quad (\text{negative helicity}) \tag{11.2.54}$$

$$\gamma^{+}(q) = \left[\frac{M_s}{2q}\right]^2 \mathrm{Im}\,\Sigma^{+}(q) \quad ; \quad (\text{positive helicity}) \tag{11.2.55}$$

$$\Delta^{-}(q) = \mathrm{Re}\;\Sigma^{-}(q) \quad ; \quad (\text{negative helicity}) \tag{11.2.56}$$

$$\Delta^{+}(q) = \left[\frac{M_s}{2q}\right]^2 \mathrm{Re}\,\Sigma^{+}(q) \quad ; \quad (\text{positive helicity}) \tag{11.2.57}$$

The damping rates for negative $(-)$ and positive $(+)$ sterile- like neutrinos respectively are given by

$$\Gamma^{\mp}_{2}(q) = 2\,\Big(\theta^{\mp}_{eff}(q)\Big)^2\,\gamma^{\mp}(q) \tag{11.2.58}$$

as discussed in the next section these rates determine the production rates of sterile neutrinos of each helicity. For the active-like neutrinos we find

$$\Gamma_1^-(q) = 2\left[1-\left(\theta_{eff}^-(q)\right)^2\right]\ \mathrm{Im}\,\Sigma^-(q) \tag{11.2.59}$$

$$\Gamma_1^+(q) = 2\left[1-\left(\theta_{eff}^-(q)\right)^2\right]\left[\frac{M_1}{2q}\right]^2\mathrm{Im}\,\Sigma^+(q)\,, \tag{11.2.60}$$

for negative $(-)$ and positive $(+)$ helicity respectively, the latter rate can be safely neglected for the light active-like neutrinos. The effective mixing angles are given by

$$\theta_{eff}^{\pm}(q) = \frac{\theta}{\left[\left(1+\frac{\Delta^{\pm}(q)}{\xi}\right)^2+\left(\frac{\gamma^{\pm}(q)}{\xi}\right)^2\right]^{1/2}}\,, \tag{11.2.61}$$

where in the relativistic limit

$$\frac{\Delta^-(q)}{\xi} = \frac{2q}{M_s^2}\,\mathrm{Re}\Sigma^-(q)\ \ ;\ \ \frac{\Delta^+(q)}{\xi} = \frac{\mathrm{Re}\Sigma^+(q)}{2q} \tag{11.2.62}$$

$$\frac{\gamma^-(q)}{\xi} = \frac{2q}{M_s^2}\,\mathrm{Im}\Sigma^-(q)\ \ ;\ \ \frac{\gamma^+(q)}{\xi} = \frac{\mathrm{Im}\Sigma^+(q)}{2q}\,, \tag{11.2.63}$$

where $\Sigma^{\pm}$ are given by (11.2.52,11.2.53).

The important observation is that the self-energy $\Sigma(\omega,q)$ is calculated in *the standard model for massless flavor neutrinos.*

The result for the effective mixing angle (11.2.61) is valid for $\theta \ll 1$ and (11.2.62,11.2.63) are valid in the relativistic limit $q \gg M_s$ but otherwise *general* and *valid to all orders in perturbation theory in standard model couplings.*

A Mikheyev-Smirnov-Wolfenstein[30] (MSW) resonance is available whenever

$$\frac{\Delta^{\pm}(q)}{\xi} = -1\,, \tag{11.2.64}$$

for $\cos(\theta) \simeq 1$. However, this resonance is *screened* by the imaginary part (damping rate), and under the condition that $\theta \ll 1$ and the validity of the approximations leading to the above results, the possible presence of this resonance will not yield a large enhancement in the effective mixing angle. This aspect will be discussed in detail in section (11.4).

These expressions are one of the main results of this study and summarize the effective mixing angles and damping rates *generically* for standard model interactions under the assumptions of the validity of the relativistic approximation, perturbative expansion and $\theta \ll 1$.

From the expressions (11.2.28,11.2.29)) and (11.2.34,11.2.35) one finds that for $\theta \ll 1$

$$s^h(\omega = q, q) \simeq \theta^h_{eff}(q)\, e^{-i\phi^h(q)} \quad ; \quad \phi^h(q) = \tan^{-1}\left[\frac{\gamma^h(q)}{\xi + \Delta^h(q)}\right], \tag{11.2.65}$$

where the phase is irrelevant for transition probabilities and the quantum kinetic description of sterile neutrino production (transition probabilities per unit time see below).

11.3 QUANTUM KINETICS: PRODUCTION RATES

In order to understand how to extract the rate of sterile-like neutrino production from the damping rate obtained from the self energy and the effective mixing angle in the medium, let us consider first the quantum kinetics of production of the sterile-like mass eigenstate from W-decay in the case of vacuum mixing angle. This analysis clearly shows how the mixing angle in the medium enters in the quantum kinetic equation with a straightforward generalization to more general production processes.

If the mass M_2 of the heavy, sterile-like neutrino is such that $M_w > M_2 + m_l$ with m_l the mass of the charged lepton l, then the mass eigenstate corresponding to the sterile-like neutrino can be produced from W-decay, a similar argument applies to Z-decay if $M_z > M_2 + M_1$. The charged current interaction vertex for the case of one generation is

$$\mathcal{L}_{cc} = \frac{g_w}{\sqrt{2}}\left[\bar{l}\gamma^\mu \mathbb{L}\ \nu_l\ W_\mu + h.c.\right]. \tag{11.3.1}$$

Writing the flavor eigenstate ν_l in the mass basis as

$$\nu_l = \cos(\theta)\nu_1 + \sin(\theta)\nu_2 \tag{11.3.2}$$

with ν_1 being the active-like and ν_2 the sterile-like mass eigenstates with M_1, M_2 respectively, yielding an interaction vertex for the sterile-like mass eigenstate ν_2

$$\mathcal{L}_{s,cc} = \frac{g_w}{\sqrt{2}} \sin(\theta) \Big[\bar{l} \gamma^\mu \mathbb{L} \; \nu_2 \; W_\mu + h.c. \Big] \,. \tag{11.3.3}$$

The dynamics of the production of ν_2 from the process $W \to \bar{l} \; \nu_2$ is obtained via the quantum kinetic equation for the process $W \rightleftharpoons \bar{l} \nu_2$[447], namely

$$\frac{dn_2(q;t)}{dt} = \frac{dn_2(q;t)}{dt}\Big|_{\text{gain}} - \frac{dn_2(q;t)}{dt}\Big|_{\text{loss}} \,, \tag{11.3.4}$$

where $n_2(q;t)$ is the distribution function of the sterile-like mass eigenstate ν_2 and the gain and loss terms are extracted from the usual transition probabilities per unit time,

$$\frac{dn_2(q;t)}{dt}\Big|_{\text{gain}} = \tag{11.3.5}$$
$$\frac{2\pi \; \sin^2(\theta)}{2E_2(q)} \int \frac{d^3k \, |\mathcal{M}_{fi}|^2}{(2\pi)^3 2E_W(p) 2E_l(k)} \quad N_B(p)(1 - n_l(k))(1 - n_2(q;t)) \, \delta(E_W(p) - E_l(k) - E_2(q))$$
$$\frac{dn_2(q;t)}{dt}\Big|_{\text{loss}} = \tag{11.3.6}$$
$$\frac{2\pi \; \sin^2(\theta)}{2E_2(q)} \int \frac{d^3k \, |\mathcal{M}_{fi}|^2}{(2\pi)^3 2E_W(p) 2E_l(k)} \quad (1 + N_B(p)) n_l(k) n_2(q;t) \, \delta(E_W(p) - E_l(k) - E_2(q)) \,,$$

where $p = |\vec{k} + \vec{q}|$ and

$$N_B(p) = \frac{1}{e^{E_W(p)/T} - 1} \quad ; \quad n_l(k) = \frac{1}{e^{E_l(k)/T} + 1} \,, \tag{11.3.7}$$

and $|\mathcal{M}_{fi}|^2$ is the usual transition matrix element for $W \to \bar{l}\nu$ and we have assumed that the W vector boson and charged lepton l are in thermal equilibrium and displayed explicity the factor $\sin^2(\theta)$ factored out of the M_{fi}. Therefore the quantum kinetic equation (11.3.4) becomes of the form

$$\frac{dn_2(q;t)}{dt} = \Gamma^{<}(q)(1 - n_2(q;t)) - \Gamma^{>}(q) n_2(q;t) \tag{11.3.8}$$

where the gain and loss *rates* are

$$\Gamma^{<}(q) = \frac{2\pi \; \sin^2(\theta)}{2E_2(q)} \int \frac{d^3k \, |\mathcal{M}_{fi}|^2}{(2\pi)^3 2E_W(p) 2E_l(k)} \, N_B(p)(1 - n_l(k)) \, \delta(E_W(p) - E_l(k) - E_2(q)) \tag{11.3.9}$$

$$\Gamma^{>}(q) = \frac{2\pi\,\sin^2(\theta)}{2E_2(q)} \int \frac{d^3k\,|\mathcal{M}_{fi}|^2}{(2\pi)^3 2E_W(p) 2E_l(k)} \,(1+N_B(p))n_l(k)\,\delta(E_W(p)-E_l(k)-E_2(q))\,. \tag{11.3.10}$$

Because the W, l_α are in thermal equilibrium the gain and loss rates obey the *detailed balance condition*

$$\Gamma^{<}(q)\,e^{E_2(q)/T} = \Gamma^{>}(q)\,, \tag{11.3.11}$$

which can be confirmed straightforwardly from the explicit expressions (11.3.9,11.3.10) using the energy conserving delta functions and the relations

$$1+N_B(E) = e^{E/T}N_B(E) \;\;;\;\; 1-n_l(E) = e^{E/T}n_l(E)\,. \tag{11.3.12}$$

Using (11.3.11) the quantum kinetic equation (11.3.8) reads

$$\frac{dn_2(q;t)}{dt} = \Gamma_2(q)\Big[n_{eq}(q) - n_2(q;t)\Big]\,, \tag{11.3.13}$$

where

$$n_{eq}(q) = \frac{1}{e^{\frac{E_2(q)}{T}}+1} \tag{11.3.14}$$

is the *equilibrium* (Fermi-Dirac) distribution function and

$$\begin{aligned}\Gamma_2(q) &= \Gamma^{>}(q)+\Gamma^{<}(q) \qquad (11.3.15)\\ &= \frac{2\pi\,\sin^2(\theta)}{2E_2(q)} \int \frac{d^3k\,|\mathcal{M}_{fi}|^2}{(2\pi)^3 2E_W(p) 2E_l(k)} \Big[N_B(p)+n_l(k)\Big]\,\delta(E_W(p)-E_l(k)-E_2(q))\,.\end{aligned}$$

The approach to equilibrium is studied by writing $n_2(q;t) = n_{eq}(q) + \delta n_2(q;t)$, it follows from (11.3.8) that

$$\delta n_2(q;t) = \delta n_2(q;0)\,e^{-\Gamma_2(q)\,t}\,. \tag{11.3.16}$$

The relaxation rate $\Gamma_2(q)$ is precisely the damping rate of single (quasi) particle excitations (11.2.47) as discussed in refs.[450, 451, 421]. Neutral current interactions are treated similarly by passing to the mass basis and keeping only the linear term in $\sin(\theta) \simeq \theta$ for $\theta \ll 1$. It is clear from (11.3.16) that

$$\Gamma_2(q) = \sin^2(\theta)\,\Gamma_{sm}(q) \tag{11.3.17}$$

where $\Gamma_{sm}(q)$ is the rate calculated in the standard model for the production of a massive neutrino, furthermore, it is given by the *imaginary part of the standard model flavor neutrino*

self energy evaluated on the massive neutrino mass shell. In the limit of a relativistic sterile-like mass eigenstate, $\Gamma_{sm}(q)$ is *identical* to the imaginary part of the self-energy for an active massless neutrino in the standard model. In fact in this limit the quantum kinetic equation for the active-like mass eigenstate in the relativistic limit is the same as (11.3.13) but with $\sin^2(\theta)$ in (11.3.17) replaced by $\cos^2(\theta)$.

Fundamentally the heart of the argument is simply detailed balance, a consequence of the main assumption that the plasma degrees of freedom are in thermodynamic equilibrium: the damping rate of single quasiparticle excitations $\Gamma_2(q)$ determines the approach to equilibrium in linear response[421] and for $\theta \ll 1$ the quantum kinetic equation is linear in the population n_2 to leading order in θ, therefore the gain term in the quantum kinetic equation is simply related to the relaxation rate by detailed balance. This argument is general for $\theta \ll 1$. Therefore, comparing with the damping rate for the sterile-like mass eigenstate (11.2.48), this analysis makes clear that for $\theta \ll 1$ the medium effects on the mixing angle in the quantum kinetic equation are incorporated by the simple replacement $\sin(\theta) \to \theta_{eff}(q)$ in (11.3.16,11.3.17), in other words the *full quantum kinetic equation for sterile-like production*

$$\frac{dn_2^h(q;t)}{dt} = \Gamma_2^h(q)\Big[n_{eq}(q) - n_2^h(q;t)\Big]\,, \tag{11.3.18}$$

where $\Gamma_2^{\mp}(q)$ are given by (11.2.58) with (11.2.54,11.2.54). Hence, the *production* rate of sterile-like neutrinos is

$$\Gamma^h_{prod}(q) = \Gamma_2^h(q)\, n_{eq}(q)\,. \tag{11.3.19}$$

In summary, the production rates for sterile-like neutrinos of negative $(-)$ and positive $(+\,)$helicities are given by

$$\Gamma^-_{prod}(q) = 2\Big(\theta^-_{eff}(q)\Big)^2\, \mathrm{Im}\Sigma^-(q)\, n_{eq}(q) \tag{11.3.20}$$

$$\Gamma^+_{prod}(q) = 2\Big(\theta^+_{eff}(q)\Big)^2 \left[\frac{M_s}{2q}\right]^2 \mathrm{Im}\Sigma^+(q)\, n_{eq}(q) \tag{11.3.21}$$

where the mixing angles $\theta^{\mp}_{eff}(q)$ is given by (11.2.43) with the definitions (11.2.54-11.2.57). In the production rates (11.3.20,11.3.21) $\Sigma(q)$ is the *standard model self-energy for flavor*

neutrinos evaluated on the relativistic mass shell, and n_2^h refer to the population of the *sterile-like mass eigenstate* of helicity h. Because θ_{eff} depends on helicity the matrix elements M_{fi} *should not* be averaged over helicity (spin) states.

From the expression (11.3.16) we can also glean how the helicity suppression is manifest in the case of massive neutrinos. For this it is convenient to look at the positive frequency solutions of the massive Dirac equation in the chiral representation ($\gamma^5 = diag(\mathbb{1}, -\mathbb{1})$) and in the helicity basis:

$$\mathcal{U}_+(\vec{q}) \;=\; N\begin{pmatrix} v_+(\vec{q}) \\ -\varepsilon(q)\, v_+(\vec{q}) \end{pmatrix} \;\; ; \;\; \mathcal{U}_-(\vec{q}) = N\begin{pmatrix} -\varepsilon(q)\, v_-(\vec{q}) \\ v_-(\vec{q}) \end{pmatrix} \tag{11.3.22}$$

where $v_\pm(\vec{q})$ are helicity eigenvectors (Weyl spinors) for $h = \pm 1$, and

$$N = \sqrt{E_s(q) + q} \;\; ; \;\; \varepsilon(q) = \frac{M_s}{E(q) + q}\,, \tag{11.3.23}$$

then

$$\mathbb{L}\, \mathcal{U}_+(\vec{q}) = \varepsilon(q)\; N\begin{pmatrix} 0 \\ v_+(\vec{q}) \end{pmatrix} \tag{11.3.24}$$

in the relativistic limit $q \gg M_s$, $\varepsilon \simeq M_s/2q$, this projected wave function enters in the matrix element M_{fi} for a massive positive helicity neutrino in the final state, therefore

$$|M_{fi}|^2 \propto \left(\frac{M_s}{2q}\right)^2 \tag{11.3.25}$$

in agreement with the helicity suppression for the damping rate discussed in the previous section.

Generality: Although in the above discussion we focused on the production process $W \to \overline{l}\nu_2$, the result (11.3.18) is general for $\theta \ll 1$. Consider the standard model charged and neutral current vertices, writing these in the basis of mass eigenstates the charged current vertex is linear in the mass eigenstate ν_2 therefore the vertex is $\propto \theta$. The neutral current vertex would feature a term linear in θ ($\propto \overline{\nu}_1\, \nu_2$) and another $\propto \theta^2$ ($\propto \overline{\nu}_2\, \nu_2$), for $\theta \ll 1$ this last term can be neglected and both charged and neutral current vertices are *linear* in θ and ν_2. Furthermore $\theta \ll 1$ justifies taking the "active-like" mass eigenstate to be in local thermal equilibrium (LTE) in the medium for $T \geq 0.1\,\mathrm{MeV}$ as its relaxation rate is much

larger than that of the "sterile-like" eigenstate which is suppressed by $\propto \theta^2 \ll 1$. Because the interaction vertices are linear in the neutrino field to leading order in θ, the quantum kinetic equation (gain - loss) is obviously of the form (11.3.8) and because the degrees of freedom that lead to the gain and loss terms are all in (LTE) the gain ($\Gamma^<$) and loss ($\Gamma^>$) rates must obey the detailed balance condition (11.3.11). This analysis leads directly to the quantum kinetic equation (11.3.13) after replacing $\theta \to \theta_{eff}(q)$ where q is the momentum of the sterile-like neutrino on its mass shell, the $|M_{fi}|^2$ matrix element for the gain and loss transition rates are insensitive to the phase in (11.2.65). Analyzing the approach to equilibrium leads to the identification of Γ_2 with the damping rate of the sterile-like mass eigenstate. This argument is general and the analysis presented above for $W \to \bar{l}\nu_2$ provides a direct example, which will be the focus of a detailed analysis in the next section.

11.4 STERILE PRODUCTION FROM VECTOR BOSON DECAY.

We now focus on the description of sterile(like) neutrino production via vector boson decay $W \to \bar{l}\,\nu_2 \;\;;\;\; Z^0 \to \overline{\nu}_1\nu_2$ at temperature $T \simeq M_{w,z}$, this is the leading order production process at this temperature. This temperature scale is sufficiently lower than the electroweak crossover scale $T \simeq 160\,\text{GeV}$ that the Higgs field is near its vacuum expectation value and the finite temperature corrections to the W, Z masses can be safely neglected[448]. At high temperature, the propagator of charged leptons receives substantial hard thermal loop corrections from electromagnetic interactions (and quarks from both photons and gluons) for momenta $\leq eT$[418, 419, 420, 417, 421]. However, the decay of a vector boson *at rest* in the plasma yields particles with momenta $\simeq M_{w,z}/2$, therefore the typical momenta of the charged lepton is $\mathcal{O}(T)$ and in this regime the hard thermal loop corrections are subleading and will be neglected in the following analysis, by the same reason a sterile neutrino of mass $M_2 \ll M_{w,z}$ will be produced relativistically. Low momentum sterile neutrinos (and charged leptons) *can* be produced for highly boosted vector bosons in the medium, but those excitations will be Boltzmann suppressed for $T \simeq M_{w,z}$.

We will take the charged lepton and the active-like neutrino to be massless and refer

generically to the vector boson mass as M, adapting the general result to the W, Z a posteriori. Under this approximation (justified for $M_2 \ll T \simeq M_{z,w}$) the one-loop self energy is the same for both charged and neutral current interactions, in the latter case the loop in fig. (70-(b)) includes the active-like neutrino (self-consistently) assumed to be in LTE.

The one-loop tadpole contribution from neutral currents (fig. (70)) is given by[415, 416]

$$\begin{aligned}\Sigma^t &= -\gamma^0\,\pi\,\frac{\alpha_w}{M_w^2}\sum_f C_v^{(f)}\int\frac{d^3k}{(2\pi)^3}\left[n_f(k)-\overline{n}_f(k)\right]\\ &= -\gamma^0\,\frac{\pi}{6}\,\alpha_w\left(\frac{T}{M_w}\right)^2\sum_f C_v^{(f)}\,\mu^f\left[1+\frac{\mu_f^2}{\pi^2\,T^2}\right], \qquad (11.4.1)\end{aligned}$$

where f are all the ultrarelativistic fermionic species in thermal equilibrium at temperature T and chemical equilibria with chemical potentials μ^f respectively. The tadpole Σ^t is independent of frequency and momentum and contributes only to A in (11.2.14,11.2.16). Although we quote this result as part of the general formulation, we will neglect the lepton and quark asymmetries in the following analysis setting $\mu^f = 0$ for all fermionic species, thereby neglecting the contribution Σ^t to the self-energy.

We obtain the imaginary part of the self-energy (for both helicities) in (11.2.12) from which we will obtain the real part from the dispersion relation (11.2.13).

For both charged and neutral current contributions (fig. (70,(b),(c)) for relativistic leptons, the imaginary part of the self energy is given by[416]

$$\begin{aligned}Im\Sigma(q_0,\vec{q}) &= \pi g^2\int\frac{d^3k}{(2\pi)^3}\int_{-\infty}^{\infty}dk_0\Big[1-n_f(k_0)+N_b(p_0)\Big]\gamma^\mu\rho_f(k_0,\vec{k})\rho_b(p_0,\vec{p})\gamma^\nu\,P_{\mu\nu}(p_0,\vec{p})\\ &\quad (p^\mu = q^\mu - k^\mu)\,, \qquad (11.4.2)\end{aligned}$$

where f stands for the fermionic species, either a charged lepton l for the charged current or the active neutrino ν_a (assumed in thermal equilibrium) for the neutral current contributions

and b for either vector boson in the intermediate state. The couplings and masses for the charged and neutral current contributions are

$$g \quad = \quad \begin{cases} \frac{g_w}{\sqrt{2}} & \text{CC} \\ \frac{g_w}{2\cos(\theta_w)} & \text{NC} \end{cases} \quad ; \quad M = \begin{cases} M_w & \text{CC} \\ M_z = \frac{M_w}{\cos(\theta_w)} & \text{NC} \end{cases}$$

$$\sin^2(\theta_w) \simeq 0.23 \quad ; \quad \alpha_w = \frac{g_w^2}{4\pi} \simeq \frac{1}{32} \tag{11.4.3}$$

The spectral densities are respectively (for massless fermions)

$$\rho_f(k_0, \vec{k}) = \frac{\not{k}}{2k}\Big[\delta(k_0 - k) - \delta(k_0 + k)\Big] \quad ; \quad \not{k} = \gamma^0 k^0 - \vec{\gamma}\cdot\vec{k}\,, \tag{11.4.4}$$

$$\rho_b(p_0, \vec{p}) = \frac{1}{2W_p}\Big[\delta(p^0 - W_p) - \delta(p^0 + W_p)\Big] \quad ; \quad W_p = \sqrt{p^2 + M^2} \quad ; \quad p^\mu = q^\mu - k^\mu\,, \tag{11.4.5}$$

and the projection operator

$$P_{\mu\nu}(p_0, \vec{p}) = -\left[g_{\mu\nu} - \frac{p_\mu p_\nu}{M^2}\right] \quad ; \quad p^\mu = (p^0, \vec{p}) \; ; \; M^2 \equiv M_{z,w}^2 \tag{11.4.6}$$

and

$$n_f(k_0) = \frac{1}{e^{k_0/T} + 1} \quad ; \quad N_b(p_0) = \frac{1}{e^{p_0/T} - 1}\,. \tag{11.4.7}$$

As per the discussion in the previous sections (see eqns.(11.2.26, 11.2.39)), we need the combinations $A(q_0, q) \pm B(q_0, q)$ which are obtained from the traces (11.2.52,11.2.53). We find

$$\mathrm{Im}\Big[A(q_0, q) \mp B(q_0, q)\Big] \quad = \tag{11.4.8}$$

$$\pi g^2 \int_{-\infty}^{\infty} dk_0 \int \frac{d^3k}{(2\pi)^3\; 4k\, W_p} \qquad L^{\mu\nu}\Big[Q^\pm; k\Big] P_{\mu\nu}[p]\; \overline{\rho}_f(k_0, k)\overline{\rho}_b(p_0, p)\Big[1 - n_f(k_0) + N_b(p_0)\Big]\,,$$

where

$$\overline{\rho}_f(k_0, k) \quad = \quad \Big[\delta(k_0 - k) - \delta(k_0 + k)\Big] \tag{11.4.9}$$

$$\overline{\rho}_b(p_0, p) \quad = \quad \Big[\delta(p_0 - W_p) - \delta(p_0 + W_p)\Big] \quad ; \quad p^\mu = q^\mu - k^\mu \tag{11.4.10}$$

and

$$L^{\mu\nu}\Big[Q; k\Big] = \Big[Q^\mu k^\nu + Q^\nu k^\mu - g^{\mu\nu} Q\cdot k\Big]\,. \tag{11.4.11}$$

Using the various delta functions from $\overline{\rho}_f; \overline{\rho}_b$ we find for the negative helicity component

$$L^{\mu\nu}\Big[Q^+,k\Big]P_{\mu\nu}\big[p\big] = -\frac{M^2}{q}\Big[F_1(q_0,q) + k^0\,F_2(q_0,q)\Big]\,, \tag{11.4.12}$$

with

$$\begin{aligned} F_1(q_0,q) &= \left[1-\left(\frac{(q^0)^2-q^2}{M^2}\right)\right]\left[1-\frac{\Big(q^0-q\Big)^2}{2M^2}\right] &(11.4.13)\\ F_2(q_0,q) &= 2\frac{(q^0-q)}{M^2}\left[1-\left(\frac{(q^0)^2-q^2}{2M^2}\right)\right]. &(11.4.14)\end{aligned}$$

Similarly, for the positive helicity component

$$L^{\mu\nu}\Big[Q^-,k\Big]P_{\mu\nu}\big[p\big] = \frac{M^2}{q}\Big[G_1(q_0,q) + k^0\,G_2(q_0,q)\Big]\,, \tag{11.4.15}$$

with

$$\begin{aligned} G_1(q_0,q) &= \left[1-\left(\frac{(q^0)^2-q^2}{M^2}\right)\right]\left[1-\frac{\Big(q^0+q\Big)^2}{2M^2}\right] &(11.4.16)\\ G_2(q_0,q) &= 2\frac{(q^0+q)}{M^2}\left[1-\left(\frac{(q^0)^2-q^2}{2M^2}\right)\right]. &(11.4.17)\end{aligned}$$

Note the relation

$$G_1(q^0,q) = F_1(q^0,-q) \;\;;\;\; G_2(q^0,q) = F_2(q^0,-q)\,. \tag{11.4.18}$$

Using the results above, it is straightforward to show that

$$\mathrm{Im}\Big[A(q_0,q)+B(q_0,q)\Big] = \mathrm{Im}\Big[A(-q_0,q)-B(-q_0,q)\Big]\,. \tag{11.4.19}$$

This identity relates the imaginary parts for positive energy, negative helicity neutrinos to negative energy positive helicity (anti-neutrinos) (in absence of a chemical potential). This identity guarantees that the production rate for negative (positive) helicity neutrinos is the same as for positive (negative) helicity anti-neutrinos and is a consequence of the vanishing chemical potentials under the assumption of vanishing lepton and baryon asymmetry and vanishing of the neutral current tadpole contribution.

It is convenient to change integration variables, with

$$W \equiv W_p = \sqrt{q^2 + k^2 + M^2 - 2qk\cos(\varphi)} \Rightarrow \frac{dW}{d\cos(\varphi)} = -\frac{qk}{W_p} \tag{11.4.20}$$

therefore

$$\frac{d^3k}{W_p} = (2\pi)k^2 dk \frac{d(\cos(\varphi)}{W_p} = -(2\pi)\frac{kdk\, dW}{q}\,, \tag{11.4.21}$$

yielding

$$\begin{aligned} \mathrm{Im}\Big[A(q_0,q) \mp B(q_0,q)\Big] &= \frac{g^2}{16\pi\, q}\int_{-\infty}^{\infty} dk_0 \int_0^{\infty} dk \int_{W_-}^{W+} dW \\ L\Big[Q^{\pm};k\Big] &\cdot P\big[p\big]\ \overline{\rho}_f(k_0,k)\overline{\rho}_b(p_0,p)\Big[1 - n_f(k_0) + N_b(p_0)\Big]\,, \end{aligned} \tag{11.4.22}$$

where $p^0 = q^0 - k^0$. Now in terms of the integration variables k^0, k, W

$$\overline{\rho}_b = \Big[\delta(q^0 - k^0 - W) - \delta(q^0 - k^0 + W)\Big] \tag{11.4.23}$$

and the integration limits in W are

$$W^{\pm} = \sqrt{(q \pm k)^2 + M^2}\,. \tag{11.4.24}$$

The technical details of the calculation of the spectral densities is relegated to appendix (11.10.1), we neglect the zero temperature contribution focusing solely on the finite temperature terms.

11.4.1 Imaginary parts (damping rates)

We can now obtain the imaginary parts evaluated on the relativistic mass shells $q_0 \simeq q$ (for positive energy neutrinos). The analysis of the support for the delta functions in the appendix shows that on the relativistic mass shell $q_0 = q$ the only contribution to the imaginary parts arises from the product (11.10.13 with $k^+ = \infty, k^- = M^2/4q$)

$$-\delta(k^0 + k)\delta(q - k^0 - W_p) \quad ; \quad \vec{p} = \vec{q} + \vec{k}$$

corresponding to the process $W \to \bar{l}\nu_2$ (the anti lepton $\bar{l}$ is recognized in the delta function $\delta(k^0 + k)$ which determines that the energy is $-k$).

This contribution to (11.4.9) yields

$$\begin{aligned} &\mathrm{Im}\Big[A(q,q) \mp B(q,q)\Big] \quad = \qquad\qquad (11.4.25)\\ &-\pi g^2 \int \frac{d^3k}{(2\pi)^3\ 4k\,W_p} \quad L^{\mu\nu} \quad \Big[Q^{\pm};k\Big] P_{\mu\nu}\big[p\big] \quad \Big[n_f(k) + N_b(W_p)\Big]\,\delta(W_p - q - k)\,, \end{aligned}$$

which is *precisely* the expression for the rate Γ_2 in the quantum kinetic equation (11.3.16) with ultrarelativistic neutrinos and charged leptons[1](up to the prefactor $\sin^2(\theta)$). The helicity suppression factor arises similarly to the discussion after (11.3.17).

For negative helicity the terms $F_1(q^0 = q, q) = 1, F_2(q^0 = q, q) = 0$ and with the definitions (11.2.51) we find for negative $(-)$ and positive $(+)$ helicities respectively

$$\mathrm{Im}\Sigma^-(q) = \frac{g^2 T}{16\pi}\frac{M^2}{q^2}\ln\left[\frac{1 + e^{-M^2/4qT}}{1 - e^{-M^2/4qT}\,e^{-q/T}}\right] \qquad (11.4.26)$$

$$\mathrm{Im}\Sigma^+(q) = \frac{g^2 T}{16\pi}\left\{\ln\left[\frac{1 + e^{-M^2/4qT}}{1 - e^{-M^2/4qT}\,e^{-q/T}}\right] + \frac{2T}{q}\sum_{n=1}^{\infty}\frac{e^{-n\,M^2/4qT}}{n^2}\left(e^{-n\,q/T} - (-1)^n\right)\right\}. \qquad (11.4.27)$$

These expressions clearly show the suppression for $q \ll M$ for $M \simeq T$ as a consequence of the fact that the decay products feature energy $\simeq M/2$. These results pertain generically to a vector boson of mass M, we must add the contributions from the charged and neutral vector

[1]The lepton tensor $L_{\mu\nu}$ is in terms of $Q^{\pm}$ that is divided by the energy of the relativistic neutrino (see the definitions (11.2.50)).

bosons with their respective masses and couplings. Anticipating the study with cosmological expansion in the next sections we take as a reference mass that of the W vector boson M_w and introduce the dimensionless variables

$$\tau \equiv \frac{M_w}{T} \quad ; \quad y = \frac{q}{T} \tag{11.4.28}$$

with the standard model relations (11.4.3) and defining

$$\begin{aligned} L[\tau,y] &= \ln\left[\frac{1+e^{-\tau^2/4y}}{1-e^{-\tau^2/4y}\,e^{-y}}\right] \\ \sigma[\tau,y] &= \frac{2}{y}\sum_{n=1}^{\infty}\frac{e^{-n\,\tau^2/4y}}{n^2}\left(e^{-n\,y}-(-1)^n\right) \quad ; \quad c \equiv \cos(\theta_w) \simeq 0.88 \end{aligned} \tag{11.4.29}$$

the sum of the contributions yield for $\gamma^{\mp}(q)$ (11.2.54,11.2.55)

$$\gamma^{-}(\tau;y) = M_w\,\frac{\alpha_w\,\tau}{y^2}\left[\frac{1}{8}\,L[\tau,y]+\frac{1}{16c^4}\,L\big[\frac{\tau}{c},y\big]\right] \tag{11.4.30}$$

$$\gamma^{+}(\tau;y) = \alpha_w\,M_w\Big(\frac{M_s}{M_w}\Big)^2\,\frac{\tau}{4y^2}\left\{\frac{1}{8}\Big(L[\tau,y]+\sigma(\tau,y)\Big)+\frac{1}{16c^2}\Big(L\big[\frac{\tau}{c},y\big]+\sigma\big[\frac{\tau}{c},y\big]\Big)\right\}. \tag{11.4.31}$$

The helicity suppression of the positive helicity rate $\gamma^{+}(q)$ is manifest in the ratio M_s^2/M_w^2, this is expected on the grounds that the typical momentum of the emitted neutrino is $\simeq M_w/2$. As a function of $y = q/T$ the rates feature a maximum at $\simeq \tau^2/8$, they are displayed in figs. (71,72).

The suppression of the imaginary parts on-shell (damping rates) as $y \to 0$ has a simple explanation: for a vector boson of mass M decaying *at rest* in the plasma into two relativistic leptons, energy conservation implies that each lepton carries a momentum $M/2$, and for $\tau \simeq 1$ this implies $y \simeq 1/2$. For the neutrino to feature $y \ll 1$ it must be that the massive vector boson is *highly boosted* in the plasma but the probability of such state is exponentially suppressed, thus resulting in an exponential suppression of low momentum neutrinos.

For the mixing angles in the medium (11.2.43) we need $\gamma^{\pm}(q)/\xi$ with ξ given by (11.2.40), namely

$$\frac{\gamma^{-}(\tau;y)}{\xi} = \Big(\frac{M_w}{M_s}\Big)^2\,I^{-}(\tau;y) \quad ; \quad I^{-}(\tau,y) = \frac{2\alpha_w}{y}\left[\frac{1}{8}\,L[\tau,y]+\frac{1}{16c^4}\,L\big[\frac{\tau}{c},y\big]\right] \tag{11.4.32}$$

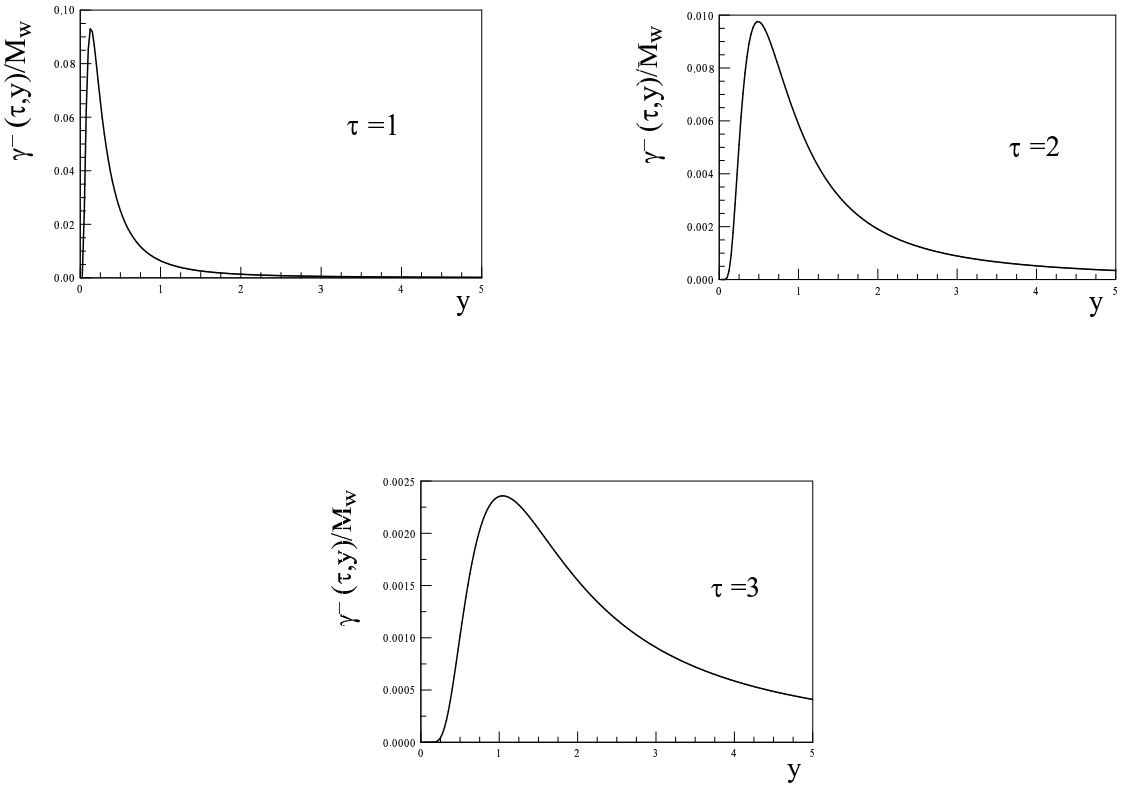


Figure 71: $\gamma^-(q)/M_w$ vs. $y = q/T$ for $\tau = M_w/T = 1, 2, 3$ respectively.

$$\frac{\gamma^+(\tau;y)}{\xi} \equiv I^+(\tau;y) = \frac{\alpha_w}{2y}\left\{\frac{1}{8}\Big(L[\tau,y] + \sigma(\tau,y)\Big) + \frac{1}{16c^2}\Big(L\big[\frac{\tau}{c},y\big] + \sigma\big[\frac{\tau}{c},y\big]\Big)\right\}. \quad (11.4.33)$$

Figs.(73,74) display $I^{\mp}(\tau, y)$ for $\tau = 1, 2, 3$. The main observation is that $I^{\mp} \ll 1$ in the whole range of y for $\tau \gtrsim 1$. This is important: note that γ^-/ξ is *enhanced* by the factor M_w^2/M_s^2 therefore for $M_w/M_s \gtrsim 10^2$ it follows that $\gamma^-/\xi \gg 1$ for $\tau \gtrsim 1$ for $y \simeq 1$, this will result in a large suppression of the effective mixing angle in the medium. On the other hand the helicity suppression implies that $\gamma^+/\xi = I^+(\tau, y) \ll 1$ in the whole range of y for $\tau \geq 1$, this will result in a vanishingly small correction to the effective mixing angle in the

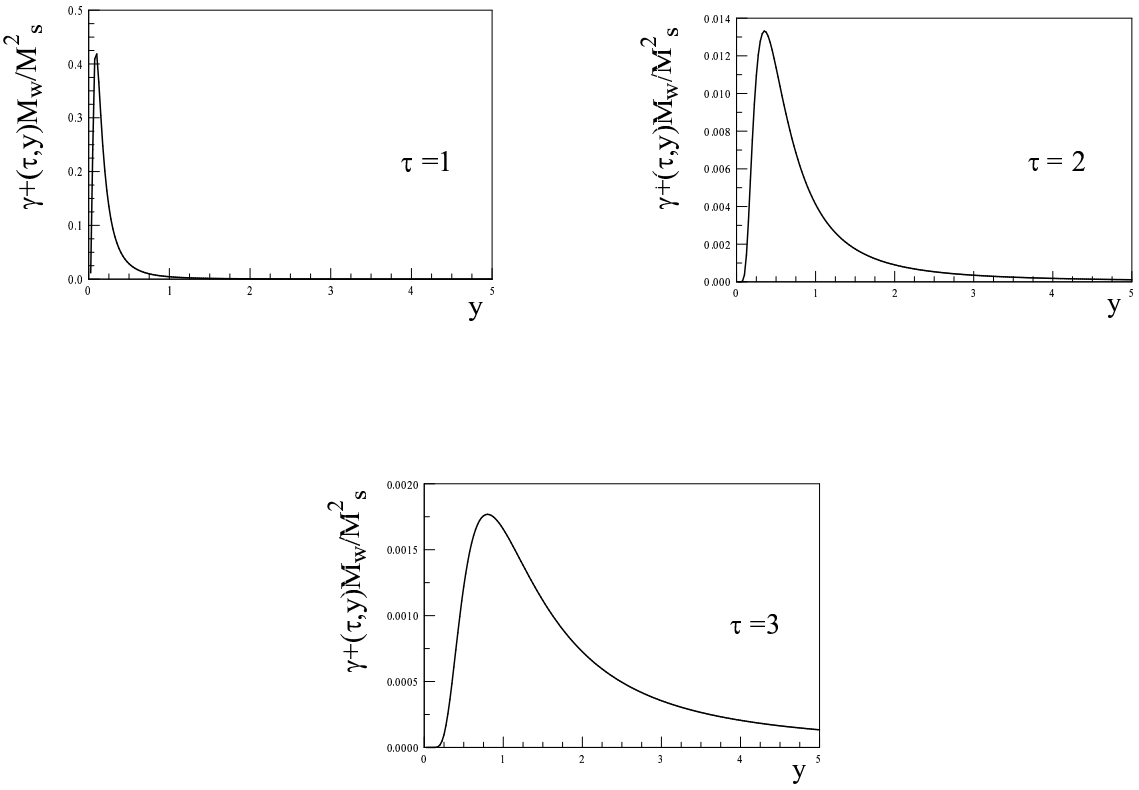


Figure 72: $\gamma^+(q)M_w/M_s^2$ vs. $y = q/T$ for $\tau = M_w/T = 1, 2, 3$ respectively.

medium, which in this case will be nearly the same as that for the vacuum. These points will be revisited again below when we discuss the corrections to the mixing angle *vis-a-vis* the production rate in the expanding cosmology.

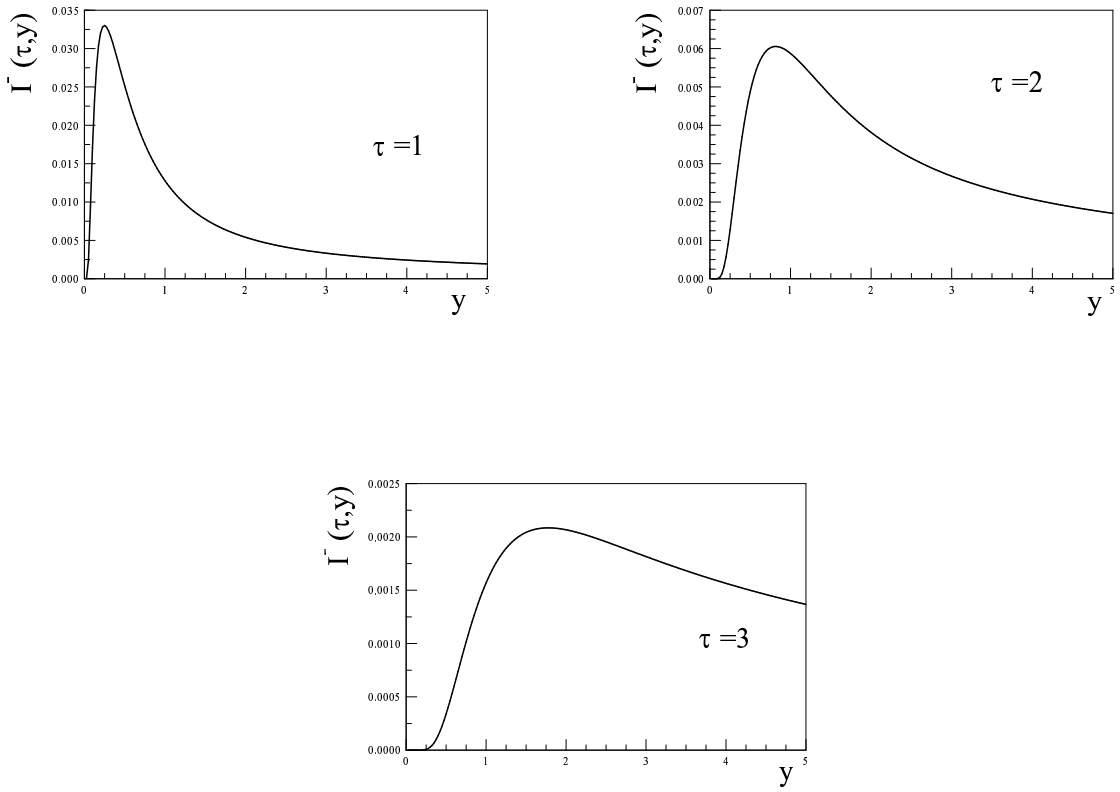


Figure 73: $I^{-}(\tau, y)$ vs. $y = q/T$ for $\tau = M_w/T = 1, 2, 3$ respectively.

11.5 REAL PART: INDEX OF REFRACTION

The index of refraction or real part of the self-energy is obtained from the dispersive representation (11.2.13). In appendix (11.10.1) we provide the details of the calculation of $\mathrm{Re}\Sigma^{\mp}$, both are of the form

$$\mathrm{Re}\Sigma^{\mp}(q) = \frac{g^2\, T}{16\pi^2}\, \frac{M^2}{q^2}\, \mathcal{K}^{\mp}[\frac{M}{T}, \frac{q}{T}]\,, \tag{11.5.1}$$

where $\mathcal{K}^{\mp}[\tau, y]$ are dimensionless functions of the ratios $M/T; y = q/T$ calculated numerically implementing the steps detailed in the appendix. Combining the contributions from charged

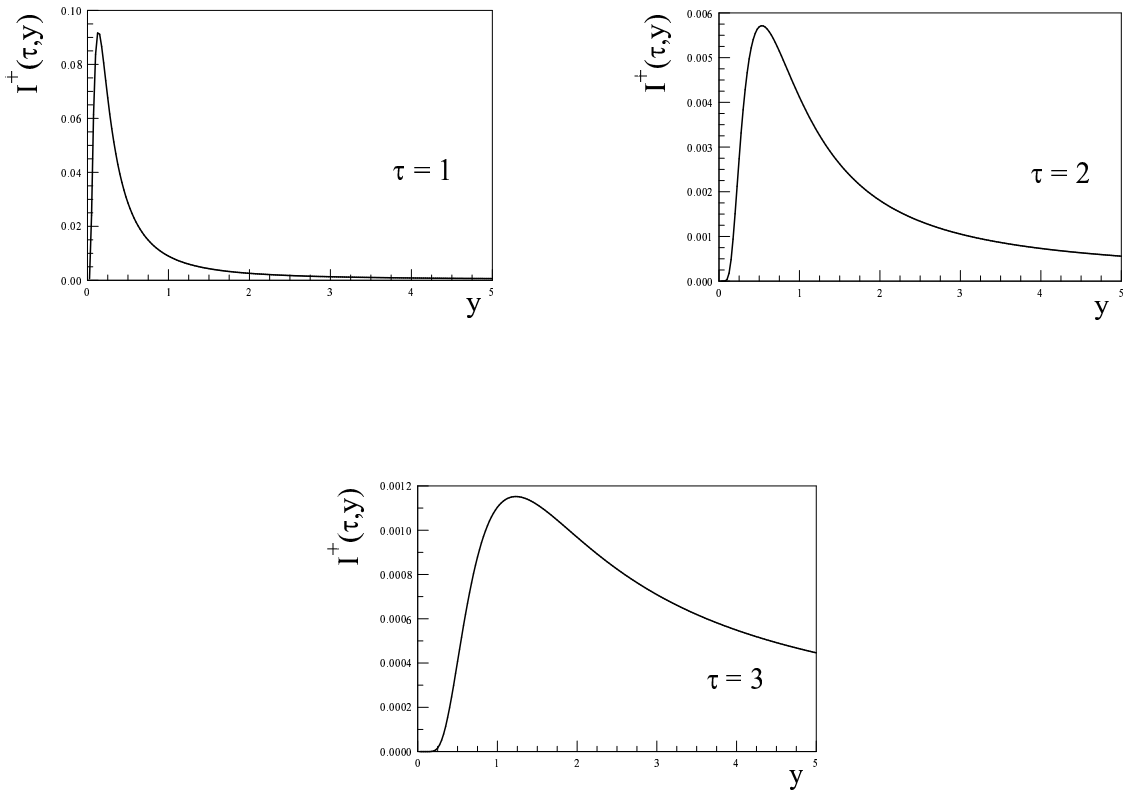


Figure 74: $I^+(\tau, y)$ vs. $y = q/T$ for $\tau = M_w/T = 1, 2, 3$ respectively.

and neutral currents we find

$$\mathrm{Re}\Sigma^{\mp}(\tau, y) = M_w \frac{\tau}{y^2} \frac{\alpha_w}{4\pi} \Big[\frac{1}{2} \mathcal{K}^{\mp}[\tau, y] + \frac{1}{4c^4} \mathcal{K}^{\mp}[\frac{\tau}{c}, y] \Big] . \tag{11.5.2}$$

Of relevance for the in-medium mixing angle are the ratios $\Delta^{\pm}/\xi$ with $\Delta^{\pm}$ given by (11.2.56,11.2.57) and ξ by (11.2.40).

Negative helicity: For negative helicity, with the definitions (11.2.40,11.2.56) we find

$$\frac{\Delta^-(q)}{\xi} = \Big(\frac{M_w}{M_s} \Big)^2 \; J^-(\tau, y) \, , \tag{11.5.3}$$

where

$$J^-(\tau, y) = \frac{\alpha_w}{2\pi\, y}\left[\frac{1}{2}\,\mathcal{K}^-[\tau, y] + \frac{1}{4c^4}\,\mathcal{K}^-[\frac{\tau}{c}, y]\right]. \tag{11.5.4}$$

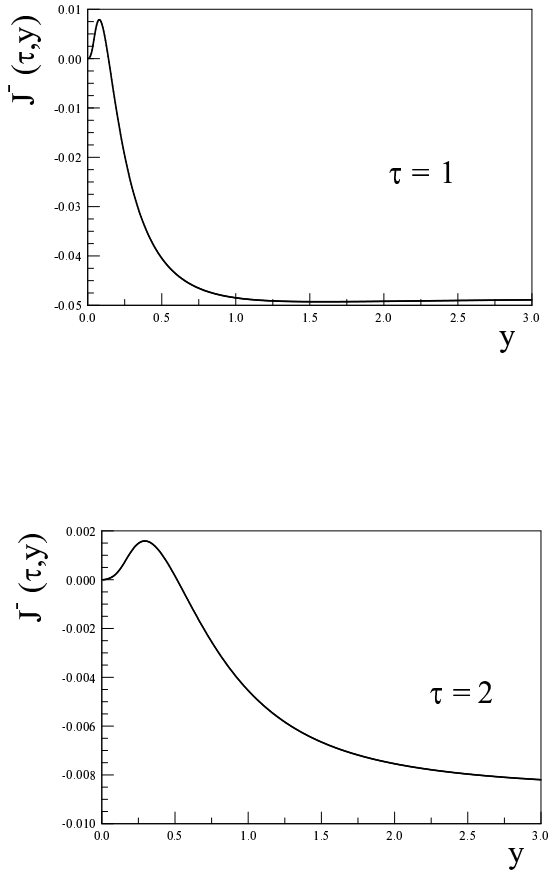


Figure 75: $J^-(\tau, y)$ vs. $y = q/T$ for $\tau = M_w/T = 1, 2$ respectively.

Low temperature limit: The limit $\tau \gg 1; \tau \gg y$ ($q \ll M_w$) affords an analytic treatment the details of which are summarized in the appendix. In this limit we find that the general form of the real part of the self-energy is given by

$$\mathrm{Re}\Sigma^-(q) = \frac{14\pi^2}{90}\, g^2 \left(\frac{T}{M}\right)^4 q \tag{11.5.5}$$

adding the charged and neutral current contributions we find in this limit

$$\mathrm{Re}\Sigma^{-}_{tot}(\tau, y) = M_w \, \frac{28\,\pi^3\,\alpha_w}{90}\Big[1 + \frac{1}{2}\,\cos^2(\theta_w)\Big]\,\frac{y}{\tau^5}\,, \tag{11.5.6}$$

which agrees with those of refs.[415, 416], and

$$\frac{\Delta^-}{\xi} = \frac{28\pi^3\,\alpha_w}{45}\left(\frac{M_w}{M_s}\right)^2\,\Big[1 + \frac{1}{2}\,\cos^2(\theta_w)\Big]\,\frac{y^2}{\tau^6}\,, \tag{11.5.7}$$

Positive Helicity For positive helicity, with the definitions (11.2.40,11.2.57) we find

$$\frac{\Delta^+(q)}{\xi} = J^+(\tau, y)\,, \tag{11.5.8}$$

where

$$J^+(\tau, y) = \frac{\alpha_w\,\tau^2}{8\pi\,y^3}\left[\frac{1}{2}\,\mathcal{K}^+[\tau, y] + \frac{1}{4c^4}\,\mathcal{K}^+[\frac{\tau}{c}, y]\right]. \tag{11.5.9}$$

Fig.(76) displays $J^+(\tau, y)$ vs. y for $\tau = 1, 2$, we see that $J^+ \ll 1$ for all values of y diminishing rapidly as a function of τ. This results in a small (and negligible) correction to the mixing angle in the medium.

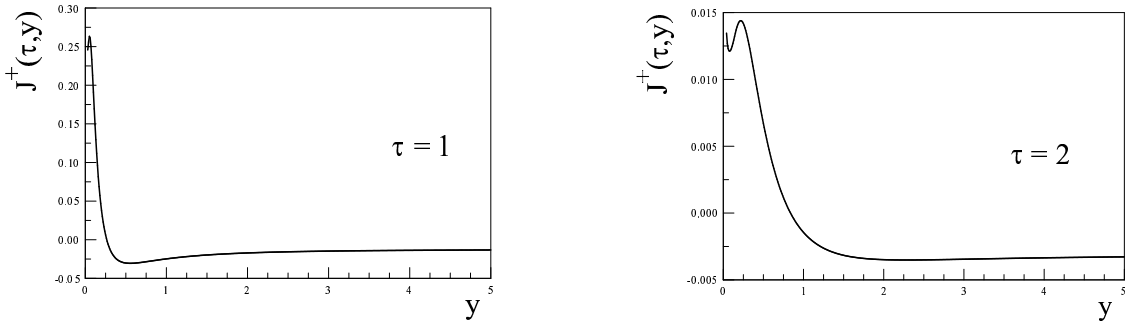


Figure 76: $J^+(\tau, y)$ vs. $y = q/T$ for $\tau = M_w/T = 1, 2$ respectively.

Low temperature limit: As in the previous case, the limit $\tau \gg 1; \tau \gg y$ ($q \ll M_w$) can be obtained analytically, see the appendix for details. The general form of the real part of the self-energy in this case is

$$\mathrm{Re}\Sigma^{+}(q) = \frac{14\pi^2}{180}\, g^2 \left(\frac{T}{M}\right)^4 q \tag{11.5.10}$$

adding the charged and neutral current contributions we find in this limit

$$\mathrm{Re}\Sigma^{+}_{tot}(\tau, y) = M_w\, \frac{28\pi^3\, \alpha_w}{180}\Big[1 + \frac{1}{2}\, \cos^2(\theta_w)\Big]\, \frac{y}{\tau^5}\,, \tag{11.5.11}$$

adding the charged and neutral current contributions and including the helicity suppression factor we find

$$\frac{\Delta^{+}}{\xi} = \frac{7\,\pi^3\, \alpha_w}{90}\, \Big[1 + \frac{1}{2}\, \cos^2(\theta_w)\Big]\, \frac{1}{\tau^4}\,. \tag{11.5.12}$$

11.6 EFFECTIVE MIXING ANGLES AND PRODUCTION RATES

The effective mixing angles are given by (11.2.61) and in the previous sections we obtained $\Delta^{\pm}/\xi$ and $\gamma^{\pm}/\xi$.

11.6.1 Negative helicity:

For negative helicity Δ^{-}/ξ is given by (11.5.3) and γ^{-}/ξ by (11.4.32), therefore

$$\theta^2_{eff}(\tau, y) = \frac{\theta^2 \left(\frac{M_s}{M_w}\right)^4}{\left[\left(\frac{M_s^2}{M_w^2} + J^{-}(\tau, y)\right)^2 + \left(I^{-}(\tau, y)\right)^2\right]}\,. \tag{11.6.1}$$

Figures (73,75) display $I^{-}(\tau, y)$, $J^{-}(\tau, y)$ as a function of y for various values of τ. These show important features: **i:)** J^{-} vanishes and becomes negative at a value of $y^{*}(\tau)$ that increases monotonically with τ. This behavior implies that for $M_w >> M_s$ the vanishing of $J^{-}(\tau, y)$ implies an MSW resonance *in absence of lepton asymmetry* for the effective mixing angle for *negative helicity*. However, this resonance is "screened" by the contribution to the

mixing angle from the imaginary part, inspection of both J^- and I^- (see figs.73,75) and an exhaustive numerical study reveal that in the broader region $0 < y^*(\tau) \lesssim y$ the imaginary part I^- yields the dominant contribution to the denominator of (11.6.1) and there is no substantial enhancement of the mixing angle as y sweeps through the resonance for any τ. In other words, the presence of the MSW resonance does not influence the effective mixing angle in a substantial manner. The numerical analysis shows that for $M_s/M_w \lesssim 10^{-2}$ the term $(M_s/M_w)^2$ in the denominator in (11.6.3) can be safely neglected and the effective mixing angle is $\propto (M_s/M_w)^4$ reflecting a strong in-medium suppression, even when the denominator becomes large because J^-, I^- are very small as shown in the figures. The smallness of the denominator is thus compensated by the large power of $M_s/M_w \ll 1$ in the numerator.

In the low temperature limit $T \ll M_w$ the effective mixing angle in the medium is given by

$$\theta_{eff}(\tau, y) = \frac{\theta}{\left[1 + \frac{28\pi^3\,\alpha_w}{45}\left(\frac{M_w}{M_s}\right)^2\left[1 + \frac{1}{2}\,\cos^2(\theta_w)\right]\frac{y^2}{\tau^6}\right]}\,, \tag{11.6.2}$$

the contribution from the imaginary part is subleading as it is suppressed by an extra power of α_w from two loop contributions to the self-energy.

The production rate (see 11.2.58,11.3.20) is

$$\Gamma^-_{prod}(\tau, y) = \frac{M_w\,\theta^2\left(\frac{M_s}{M_w}\right)^4}{y\left[e^y + 1\right]}\,\frac{\tau\, I^-(\tau, y)}{\left[\left(\frac{M_s^2}{M_w^2} + J^-(\tau, y)\right)^2 + \left(I^-(\tau, y)\right)^2\right]}\,. \tag{11.6.3}$$

Although both J^-, I^- vanish as $y \to 0$, and the effective mixing angle reaches its maximum ($\theta_{eff} \to \theta$) the imaginary part I^- vanishes exponentially as $y \to 0$ and the production rate vanishes in this limit. The effect of the MSW resonance is "screened" by the imaginary part and the resulting production rate features a peak as a function of y for fixed τ that is a result of the competition between the peak in $I^-(y, \tau)$ (see figs. (73)) and the *increase* in $(J^-)^2$.

We will analyze further the production rate in the following section within the context of cosmological expansion.

11.6.2 Positive helicity:

For positive helicity the effective mixing angle is given by

$$\theta^2_{eff}(\tau, y) = \frac{\theta^2}{\left[\Big(1 + J^+(\tau, y)\Big)^2 + \Big(I^+(\tau, y)\Big)^2\right]} \,, \tag{11.6.4}$$

and the production rate

$$\Gamma^+_{prod}(\tau, y) = \frac{M_w\,\theta^2\left(\frac{M_s}{M_w}\right)^2}{y\Big[e^y + 1\Big]}\,\frac{\tau\, I^+(\tau, y)}{\left[\Big(1 + J^+(\tau, y)\Big)^2 + \Big(I^+(\tau, y)\Big)^2\right]} \,. \tag{11.6.5}$$

From the figures (74,76) we see that $I^+, J^+ \ll 1$ in the region in which I^+ peaks, therefore we can set $J^+ \simeq 0$; $I^+ \simeq 0$ in the denominator in (11.6.5), in other words *the in-medium contribution to the mixing angle is negligible*, namely $\theta_{eff} \simeq \theta$ and we can approximate the production rate of *positive helicity* neutrinos as

$$\Gamma^+_{prod}(\tau, y) \simeq M_w\,\theta^2\left(\frac{M_s}{M_w}\right)^2\frac{\tau\, I^+(\tau, y)}{y\Big[e^y + 1\Big]} \,. \tag{11.6.6}$$

This is an important result: the helicity suppression entails a *much weaker coupling to the medium* which in turn results in negligible in-medium correction to the mixing angles and the effective mixing angle is just the vacuum mixing angle.

The positive helicity production rate is $\propto (M_s/M_w)^2$ as a consequence of helicity suppression, whereas the negative helicity production rate (11.6.3) is $\propto (M_s/M_w)^4$ as a consequence of the in-medium suppression of the mixing angle, but is enhanced by the small denominator. As it will be discussed below, there is a range of masses and temperatures for which the negative and positive helicity rates are *comparable*: positive helicity states feature helicity suppressed couplings but nearly vacuum mixing angles, whereas negative helicity states feature stronger coupling to the medium which in turn leads to strongly suppressed in-medium mixing angles, thus the competition. This aspect is studied below.

11.7 COSMOLOGICAL PRODUCTION:

We consider a spatially flat Friedmann-Robertson-Walker cosmology during a radiation dominated stage. The effect of cosmological expansion is included by replacing the momentum in Minkowski space time with the physical momentum in the expanding cosmology, namely

$$q \to \frac{q}{a(t)} \tag{11.7.1}$$

where $a(t)$ is the scale factor and now q refers to the (constant) comoving momentum. As we focus on the production during the radiation dominated era, the physical temperature is

$$T(t) = \frac{T_0}{a(t)} \tag{11.7.2}$$

where T_0 is the temperature that the plasma would have *today* as we normalize the scale factor to 1 today. We note that the variable

$$y = \frac{q(t)}{T(t)} = \frac{q}{T_0} \tag{11.7.3}$$

is a *constant* under cosmological expansion in terms of the comoving momentum and the temperature that the plasma would feature today.

In terms of the comoving wave vector q and the invariant ratio $y = q/T$ the quantum kinetic equation in the expanding cosmology reads

$$\frac{dn_2^h(q;t)}{dt} = \Gamma_2^h(q,t)\Big[n_{eq}(q) - n_2^h(q;t)\Big], \tag{11.7.4}$$

where now $\Gamma_2^h(q,t) \equiv \Gamma_2^h(\tau(t), y)$ depends on time through $\tau(t) = M_w/T(t) = M_w a(t)/T_0$,

Under the assumption of $\theta \ll 1$ and a vanishing initial population, we neglect the build up of the population and approximate the quantum kinetic equation as

$$\frac{d\, n_2^h(t)}{dt} = \Gamma_{prod}^h(\tau(t), y) \tag{11.7.5}$$

where $\Gamma_{prod}^{\mp}(\tau(t), y)$ are given by (11.6.3),11.6.5). Since the production rate depends on time through $\tau(t)$ it is convenient to use this variable in the kinetic equation, with

$$\frac{dn_2^h(t)}{dt} = \frac{dn_2^h(\tau, y)}{d\tau}\, H(t)\, \tau(t) \tag{11.7.6}$$

where during radiation domination

$$H(t) = \frac{\dot{a}(t)}{a(t)} = 1.66\, g_{eff}^{1/2}(T)\, \frac{T^2(t)}{M_{pl}} \quad ; \quad M_{pl} = 1.22 \times 10^{19}\,\text{GeV}\,. \tag{11.7.7}$$

$g_{eff}(T)$ is the effective number of relativistic degrees of freedom, $g_{eff} \simeq 100$, and varies slowly in the temperature regime $1\,\text{GeV} < T < 100\,\text{GeV}$. We will approximate $g_{eff} \simeq 100$ and constant in this temperature range anticipating that freeze-out will occur at $T_f \simeq$ few GeV.

Negative helicity

For negative helicity we find

$$\frac{dn_2^-(\tau,y)}{d\tau} \simeq \frac{0.92 \times 10^{16}\,\theta^2 \left(\frac{M_s}{M_w}\right)^4}{y\Big[e^y+1\Big]} \frac{\tau^2\, I^-(\tau,y)}{\left[\left(\frac{M_s^2}{M_w^2} + J^-(\tau,y)\right)^2 + \Big(I^-(\tau,y)\Big)^2\right]}\,. \tag{11.7.8}$$

As both J^-, I^- decrease as the temperature decreases (and τ increases) there are two competing effects: the damping rate $\propto I^-$ decreases but the effective mixing angle *increases* as a result, for a fixed value of $y = q/T$ the production rate peaks as a function of τ and falls off sharply. We write (11.7.8) as

$$\frac{dn_2^-(\tau,y)}{d\tau} \equiv 0.92 \times 10^{16}\,\theta^2 \left(\frac{M_s}{M_w}\right)^4 R^-(y,\tau)\,. \tag{11.7.9}$$

As discussed above, for $M_s/M_w < 10^{-2}$ we find numerically that $R^-(y,\tau)$ is nearly independent of M_s. The form (11.7.9) separates the suppression factor from the effective mixing angle in terms of the prefactor $\theta^2 M_s^4/M_w^4$, whereas for $M_s/M_w < 10^{-2}$ the function $R^-(y,\tau)$ is *insensitive to the value of* M_s and only depends on standard model couplings and vector boson masses. As τ increases (temperature decreases) the effective mixing angle *increases* whereas the damping rate γ^- *decreases*, therefore we expect that $R^-(y,\tau)$ for fixed values of y to feature a peak as a function τ. The analysis in the previous sections clarifies that for $\tau \gg \sqrt{y}$ the damping rate is exponentially suppressed (see 11.4.30,11.4.30) whereas the real part (index of refraction) falls off as $1/\tau^5$ (see 11.5.7) therefore the production rate is exponentially suppressed at large τ as the mixing angle grows much slower. This entails the *freeze out* of the distribution function.

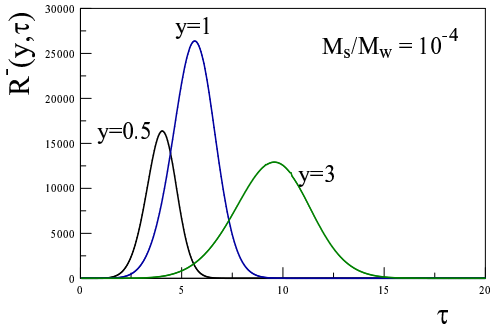


Figure 77: Rates $R^-(y,\tau)$ vs. τ for $y = 0.5, 1, 3$ respectively for $M_s/M_w = 10^{-4}$.

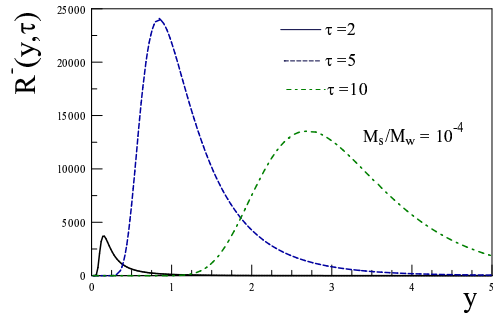


Figure 78: Rates $R^-(y,\tau)$ vs. y for $\tau = 2, 5, 10$ respectively for $M_s/M_w = 10^-4$.

This expectation is borne out from the numerical study, figs.(77,78) shows $R^-(y,\tau)$ for various values of y as a function of τ and as a function of y for $\tau = 2, 5, 10$. Numerically the case with $M_s/M_w = 10^{-4}$ is indistinguishable from that setting $M_s = 0$ in the denominator of R^- (see eqn. (11.7.8)). These figures clearly show the "freeze-out" of the distribution as

a function of τ, as the rate vanishes for large τ, larger values of y freeze-out at larger τ but with much smaller amplitudes. This feature is expected, the vector bosons are suppressed at larger temperatures and large values of y are further suppressed by the detailed balance factor $1/(e^y + 1)$. Fig. (78) shows the "filling" of the different wave-vectors: as time evolves larger y are populated but eventually larger values of y are suppressed by the Fermi-Dirac factor $n_{eq}(y)$. *Assuming* that the initial population vanishes at $\tau \simeq 1$), the asymptotic distribution function is given by

$$n_2^-(y) = \int_1^\infty \frac{dn_2^-(\tau, y)}{d\tau}\, d\tau = 0.92 \times 10^{16}\, \theta^2 \left(\frac{M_s}{M_w}\right)^4 F^-(y) \tag{11.7.10}$$

where we have defined the *frozen* distribution

$$F^-(y) = \int_1^\infty R^-(y, \tau)\, d\tau \tag{11.7.11}$$

which is shown in fig. (79). Although we have set the lower limit $\tau = 1$ for consistency in (11.7.11), we find that $R^-(y, \tau)$ vanishes as $\tau \to 0$ and the lower limit can be effectively taken to $\tau = 0$. A numerical study informs that the region $0 < y \leq 10$ which features the largest contribution to the distribution function, freezes out at $\tau_f \simeq 15$ corresponding to a freeze-out temperature for negative helicity modes $T_f^- = M_w/\tau_f \simeq 5\,\text{GeV}$.

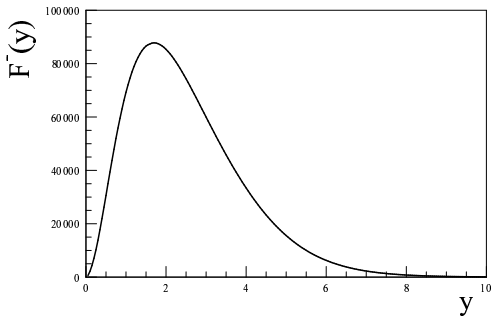


Figure 79: Asymptotic distribution function $F^-(y)$ (eqn. (11.7.11) vs. y.

The vanishing of $F^-(y)$ as $y \to 0$ is a consequence of the vanishing of the imaginary part, and a direct consequence of the decay kinematics in the medium, as explained above the $y \to 0$ is dominated by the decay of vector bosons that are highly boosted in the rest frame of the plasma, and the population of these states is highly suppressed at $T \simeq M_w$. The broadening of the distribution as compared to the damping rate (compare figures (73,79) is a consequence a longer freeze-out time resulting from the competition between a decreasing damping rate I^- and an increasing mixing angle, the modes with higher y continue to populate as the mixing angle increases but eventually as modes with large values of y are populated, their contribution is suppressed by the detailed balance factor $n_{eq}(y)$. After freeze-out, the total number density of negative helicity neutrinos produced, (equal to the total number of positive helicity antineutrinos in absence of a lepton asymmetry) is given by

$$\mathcal{N}_2^- = \frac{T^3(t)}{2\pi^2} \int_0^\infty n_2^-(y)\, y^2\, dy\,, \tag{11.7.12}$$

for which we need the result

$$\int_0^\infty y^2 F^-(y)\, dy \equiv N^- = 2.287 \times 10^6\,. \tag{11.7.13}$$

This integral is dominated by the region $0 < y \lesssim 10$, which freezes-out at $\tau \simeq 15$, with the result that

$$\frac{\int_{10}^\infty y^2 F^-(y)\, dy}{\int_0^\infty y^2 F^-(y)\, dy} = 3.9 \times 10^{-3}\,. \tag{11.7.14}$$

Normalizing the number density to that of one degree of freedom of an active massless neutrino decoupled in equilibrium at the *same* temperature, namely

$$\mathcal{N}_\nu = \frac{T^3(t)}{2\pi^2} \int_0^\infty n_{eq}(y)\, y^2\, dy\,, \tag{11.7.15}$$

where $n_{eq}(y) = 1/(e^y + 1)$, we find

$$\frac{\mathcal{N}_2^-}{\mathcal{N}_\nu} = 285\ \theta^2 \left(\frac{M_s}{\text{MeV}}\right)^4, \tag{11.7.16}$$

this ratio is constant throughout the expansion history. We note that the approximation (11.7.5) of neglecting the build-up of the population in the quantum kinetic equation (11.7.4) is consistent provided the ratio $\mathcal{N}_2^-/\mathcal{N}_\nu \ll 1$.

Positive helicity for positive helicity we find

$$\frac{dn_2^+(\tau,y)}{d\tau} \simeq 0.92 \times 10^{16}\,\theta^2 \left(\frac{M_s}{M_w}\right)^2 \frac{\tau^2\, I^+(\tau,y)}{y\Big[e^y+1\Big]}\,, \tag{11.7.17}$$

where as discussed above we have used the approximate rate (11.6.6). As for the negative helicity case it proves convenient to write (11.7.17) as

$$\frac{dn_2^+(\tau,y)}{d\tau} \simeq 0.92 \times 10^{16}\,\theta^2 \left(\frac{M_s}{M_w}\right)^2 R^+(y,\tau) \;\;;\;\; R^+(y,\tau) = \frac{\tau^2\, I^+(\tau,y)}{y\Big[e^y+1\Big]}\,, \tag{11.7.18}$$

where $R^+(\tau,y)$ is read off (11.7.17) and does not depend on M_s. Fig.(80) shows $R^+(\tau,y)$ vs. y for $\tau = 1,3,5$ respectively and fig. (81) shows $R^+(\tau,y)$ vs. τ for $y = 1,3,5$ respectively. Together these figures show the "filling" of higher momentum modes as the temperature decreases and the freeze-out of the distribution function for different wavevectors. The larger values of y take longer to be populated and freeze out later, but their contribution is strongly suppressed by the detailed balance factor $1/(e^y+1)$.

Similarly to the previous case, the asymptotic distribution function is

$$n_2^+(y) = 0.92 \times 10^{16}\,\theta^2 \left(\frac{M_s}{M_w}\right)^2 F^+(y) \;\;;\;\; F^+(y) = \int_1^\infty R^+(\tau,y)\,d\tau \tag{11.7.19}$$

The asymptotic distribution function $F_2^+(y)$ is shown in fig. (82), it is dominated by the region $0 < y \lesssim 8$ with

$$\frac{\int_8^\infty y^2 F^+(y)\,dy}{\int_0^\infty y^2 F^+(y)\,dy} = 3.3 \times 10^{-3}\,, \tag{11.7.20}$$

and freezes-out at $\tau \simeq 10$ corresponding to a freeze-out temperature $T_f^+ \simeq 8\,\mathrm{GeV}$.

We note that the distribution function for the positive helicity component is sharply peaked at small momenta $y \approx 0.5$ as compared to that for the negative helicity component which is much broader and peaks at $y \simeq 2.5$, namely the positive helicity component yields a much *colder* distribution (compare figures (79 and 82)). The reason for this discrepancy is the fact that the production rate for the negative helicity component features a competition between a diminishing damping rate, but an increasing effective mixing angle as τ increases (temperature decreases). This competition results in a longer freeze-out time allowing build-up in the population of larger momentum modes as τ evolves as discussed above. It is

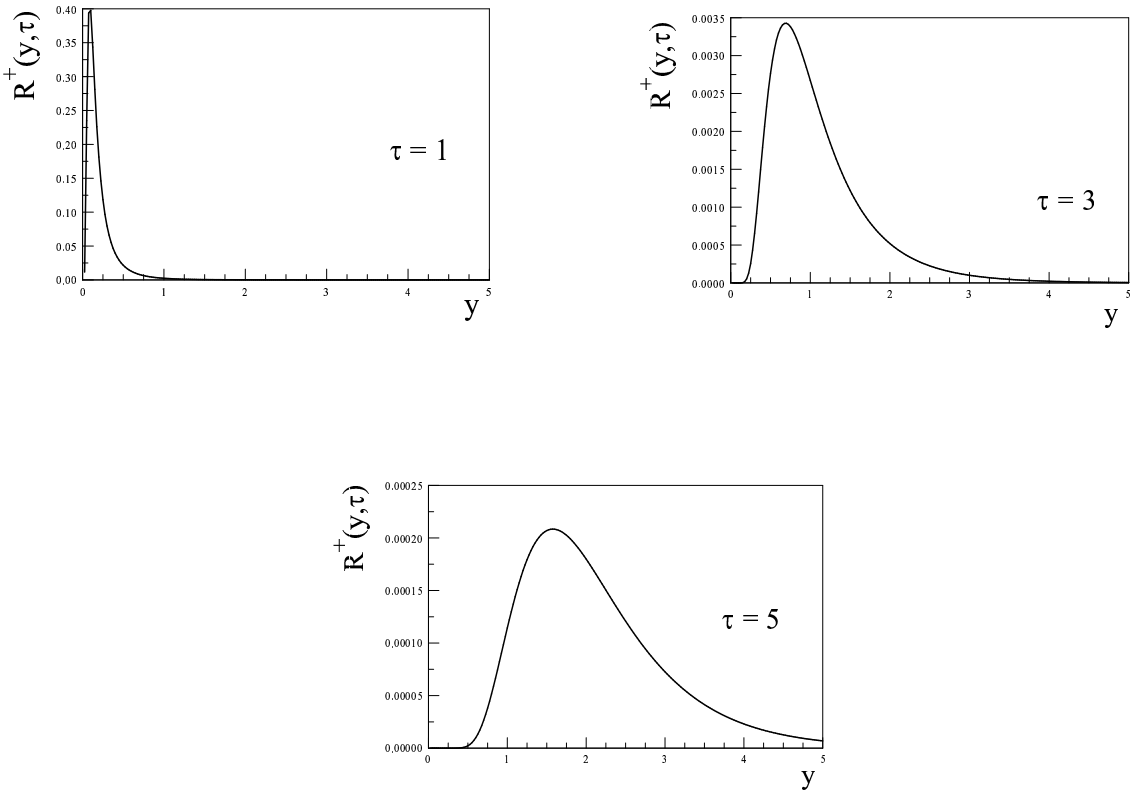


Figure 80: $R^+(\tau, y)$ vs. $y = q/T$ for $\tau = 1, 3, 5$ respectively.

remarkable that the distribution functions $F_2^{\pm}$ are very similar to those found from pion decay in ref.[447, 396], the similarity is more striking for F^-. The physical reason for this similarity is actually simply the fact that low momentum modes are suppressed since the presence of low momentum sterile neutrinos in the decay of a much more massive particle implies that this "mother" particle must be highly boosted in the plasma.

The total population at asymptotically long time is given by

$$\mathcal{N}_2^+ = \frac{1}{2\pi^2} \int_0^\infty n_2^+(y)\, y^2\, dy\,, \tag{11.7.21}$$

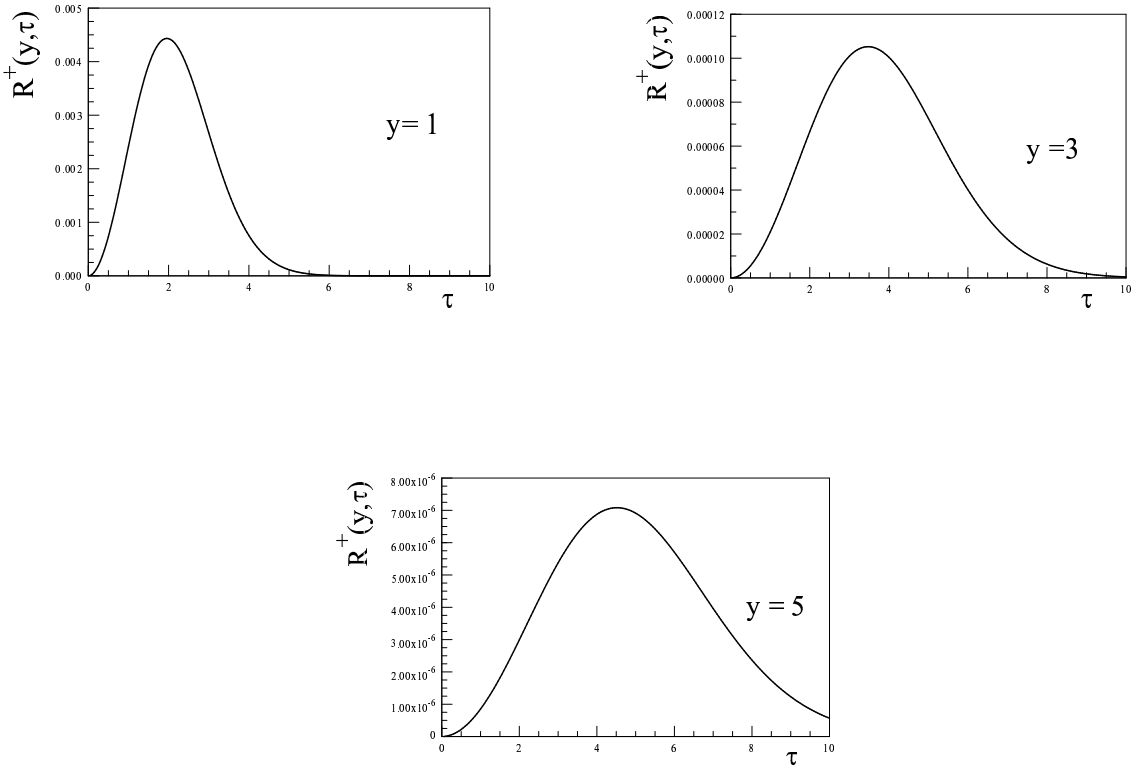


Figure 81: $R^+(y,\tau)$ vs. τ for $y = 1, 3, 5$ respectively.

which is determined by the integral

$$\int_0^\infty y^2\, F^+(y)\, dy \equiv N^+ = 0.025\,. \tag{11.7.22}$$

As in the negative helicity case, normalizing to the number density of relativistic neutrino decoupled in equilibrium at the same temperature (11.7.15) we find

$$\frac{\mathcal{N}_2^+}{\mathcal{N}_\nu} \simeq 2\times 10^4\ \theta^2\Big(\frac{M_s}{\mathrm{MeV}}\Big)^2\,. \tag{11.7.23}$$

We have studied the contributions of positive and negative helicity individually to highlight the different distribution functions and dependence on M_s, however each is simply a

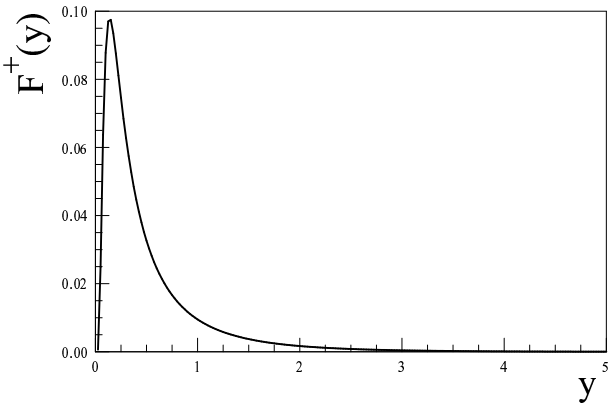


Figure 82: Asymptotic distribution function $F^+(y)$ vs. y .

different decay channel for the production of sterile-like eigenstates from the decay of vector bosons and both channels contribute to the total abundance. Hence we combine both channels to give the total density

$$\begin{aligned} n_2(y) &= n_2^+(y) + n_2^-(y) = 3.6 \left(\frac{\theta^2}{10^{-4}}\right)\left(\frac{M_s}{\mathrm{MeV}}\right)^2 f(M_s, y) \\ f(M_s, y) &= \left[\frac{F^+(y)}{N^+} + \left(\frac{M_s}{8.35\,\mathrm{MeV}}\right)^2 \frac{F^-(y)}{N^-}\right], \end{aligned} \tag{11.7.24}$$

where the normalization factors $N^{\pm}$ are given by (11.7.13,11.7.22). The effective distribution function multiplied by the phase space factor y^2 is shown in fig.(83) for $M_s = 1\,\mathrm{MeV}$ where is dominated by the *positive helicity component* and $M_s = 10\,\mathrm{MeV}$ where it is dominated by the *negative helicity component*.

This figure clearly shows the strongly non-thermal total distribution function at freeze out, it also provides a specific example of the "mixed dark matter" nature[447] when several different production channels with different kinematics and effective mixing angles contribute

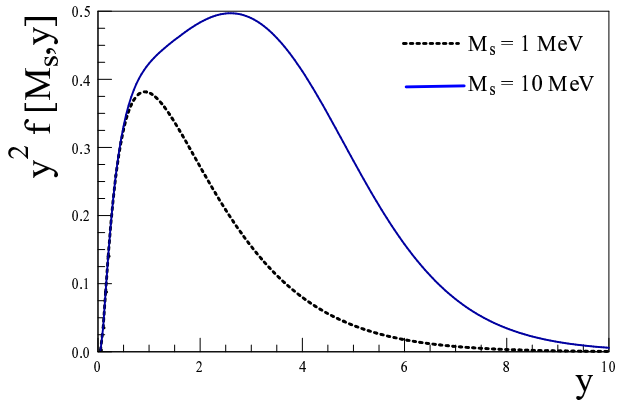


Figure 83: Total distribution function $f(M_s, y)$ multiplied by y^2 vs. y for $M_s = 1; 10\,\mathrm{MeV}$.

to the production of a sterile-like species. The "hump" in $f(M_s, y)$ for $M_s = 10\,\mathrm{MeV}$ is a result of the competition between the two channels, the negative helicity channel is hotter since its distribution is peaked at larger momenta but becomes dominant at larger M_s, whereas that of the positive helicity is colder since it is peaked at lower momenta, but dominates for smaller M_s. Accordingly, we find for the total abundance normalized to that of a single degree of freedom of a massless thermal neutrino, with $\mathcal{N}_2 = \mathcal{N}_2^+ + \mathcal{N}_2^-$

$$\frac{\mathcal{N}_2}{\mathcal{N}_\nu} \simeq 2 \left(\frac{\theta^2}{10^{-4}}\right) \left(\frac{M_s}{\mathrm{MeV}}\right)^2 \left[1 + \left(\frac{M_s}{8.35\,\mathrm{MeV}}\right)^2\right]. \tag{11.7.25}$$

The first term in the bracket is the contribution from the positive helicity states and the second from the negative helicity, both become comparable for $M_s \simeq 8.35\,\mathrm{MeV}$.

If the sterile-like neutrino is *stable*, its comoving number density would remain constant and upon becoming non-relativistic this species would contribute to dark matter a fraction given by[447]

$$\mathcal{F}_2 = \frac{\Omega_{\nu_2}\, h^2}{\Omega_{DM}\, h^2} = \frac{M_s}{7.4\,\mathrm{eV}} \left(\frac{g_{\nu_2}}{g_d}\right) \int_0^\infty n_2(y)\, y^2\, dy \tag{11.7.26}$$

where g_{ν_2} is the number of degrees of freedom for neutrinos of negative helicity, we will assume Dirac neutrinos in which case $g_{\nu_2} = 2$ accounting for neutrinos and anti-neutrinos ($g_{\nu_2} = 1$ for Majorana neutrinos) and $g_d \simeq 100$ is the number of ultrarelativistic degrees of freedom at decoupling (freeze-out) which occurs at $T_f \simeq 5 - 8\,\mathrm{GeV}$, yielding

$$\mathcal{F}_2 = 0.97 \left(\frac{\theta^2}{10^{-8}}\right) \left(\frac{M_s}{\mathrm{MeV}}\right)^3 \left[1 + \left(\frac{M_s}{8.35\,\mathrm{MeV}}\right)^2\right]. \tag{11.7.27}$$

The terms in the bracket are the contribution from the positive helicity and negative helicity respectively, the latter dominates for $M_s \gg 8.35\,\mathrm{MeV}$.

It is clear from this expression that the sterile neutrinos produced by vector boson decay *cannot* yield a substantial $\simeq$ KeV warm dark matter component, since the X-ray data constrains such component to the mass range $\simeq$ few KeV and mixing angle $\theta^2 \lesssim 10^{-10}$[393, 351, 352, 357] which according to (11.7.27) would yield a negligible abundance of such species. However, accelerator and cosmological bounds[98, 436] allow for *heavy* sterile states with masses in the MeV range and mixing angles $\lesssim 10^{-5}$, in fact these are the bounds used in the recent analysis of MeV sterile neutrinos as possible solutions to the ^7Li problem[437, 442] which we discuss further in section (11.8) below.

The results obtained above for the distribution and abundances constitute a *lower bound*, this is because we have neglected any initial population and, as it will be discussed below, we expect other processes to yield sterile-like neutrinos at various stages of the thermal history.

11.8 DISCUSSION

11.8.1 Validity of approximations

We have implemented several approximations to obtain the above results, which merit a discussion of their validity.

- Ultrarelativistic neutrinos: an obvious approximation for the active-like mass eigenstates, for the sterile-like eigenstate this implies $M_s/q \ll 1$. In the expanding cosmology this

inequality is in terms of the physical momentum $q_{phys}(t) = q/a(t)$ with q being the comoving momentum. Since $y = q_{phys}(t)/T(t)$ is a constant and $T(t) = M_w/\tau$ hence $q_{phys}(t) = yM_w/\tau$. The inequality must be evaluated at freeze-out, therefore the condition for the validity of the ultrarelativistic limit for sterile-line neutrinos is

$$\frac{yM_w}{M_s\tau_f} \gg 1,, \tag{11.8.1}$$

in the range of the distribution function with the largest support. With $\tau_f \simeq 15$ the condition (11.8.1) applies to $y \gtrsim M_s/6\,\mathrm{GeV}$, which is fulfilled for $y \gtrsim 10^{-3}$ for $M_s \simeq$ few MeV. The distribution function is exponentially suppressed at small y in both cases, therefore $M_s \simeq$ few MeV fulfills the criterion in almost the whole range but for extremely small values of y which are suppressed both by the distribution and by phase space.

- $\theta \ll 1$, this approximation was used in expanding the square roots in (11.2.28) and extracting the dispersion relations (11.2.38,11.2.42,11.2.43) and effective mixing angles (11.2.43). Assuming $\theta \ll 1$, the actual approximation is $(1 + (\Delta/\xi))^2 + (\gamma/\xi)^2 \gg \theta^2$ or in fact that $\theta_{eff} \leq \theta$ which is fulfilled in both cases. As was discussed above in the negative helicity case the (MSW) resonance when $1 + \Delta_2/\xi \simeq 0$ is actually "screened" by the term $\gamma_2/\xi = (M_w^2/M_s^2)I^-$ which is actually $\gg 1$ for $M_w/M_s > 10^2$ suppressing the effective mixing angle $\theta_{eff} \ll \theta$. Therefore this approximation is consistent, namely assuming that the vacuum mixing angle is $\ll 1$ implies that the effective mixing angle is also $\ll 1$ and the corrections are such that $\theta/\theta_{eff} \leq 1$.
- Active-like neutrinos in LTE: this approximation was invoked to obtain the neutral current contribution to the self-energy with thermalized neutrinos in the intermediate state. For $\theta \ll 1 \rightarrow \theta_{eff} \ll 1$ and $\cos(\theta_{eff}) \simeq 1$ implying that the interaction vertices of active-like neutrinos are the usual standard model ones. This, in turn implies the validity of the usual argument that leads to conclude that active neutrinos are in LTE down to $T \simeq$ MeV which is much smaller than the freeze-out temperature of sterile-like neutrinos $T_f \simeq$ few GeV. Therefore this approximation is valid all throughout the region of production via vector boson decay and even much lower temperatures down to the usual decoupling temperature $\simeq$ MeV for weak interactions.

- Perturbative expansion: the validity of perturbation theory in describing sterile-like production and freeze-out relies on *two* small dimensionless parameters: $\alpha_w \simeq 1/32$ and $\theta \ll 1$. Inspection of the ratio $\delta\omega_2^h/\omega_2(q)$ (see eqns. (11.2.38,11.2.42) clearly shows that this ratio is $\ll 1$ for $\theta_{eff} \ll 1$, $\alpha_w \ll 1$ and the ultrarelativistic limit, confirming the validity of the perturbative expansion for the description of production and freeze-out of sterile-like neutrinos.

11.8.2 Other contributions and higher orders

Production of sterile-like neutrinos from vector boson decay is the *dominant* process at $T \simeq M_{w,z}$, and is of order α_w as clearly exhibited by the results obtained above. This is the leading contribution, to the self-energy in this temperature range, namely the one-loop contributions depicted in fig. (70). In the same temperature regime there are several other processes that contribute to the imaginary part of the self-energy, hence to the production rate: heavy quark and lepton decays via charged current interactions, $q \to q\,\overline{\nu}\,\nu$, $\tau, \mu \to \overline{\nu}\,\nu\,l$ charged lepton annihilation $l^+l^- \to \overline{\nu}\nu$ (via neutral currents) and several other processes (for a more detailed discussion see[447]). These processes contribute to the imaginary part of the self-energy at *two loops*, therefore are of order α_w^2. Furthermore at $T \ll M_{w,z}$ these are further suppressed by a vector boson propagator, therefore their contribution to the imaginary part is generically $CG_F^2T^4q$ typically with $C \simeq 1$. As the temperature diminishes through the cosmological expansion, the damping rate from vector boson decay will become of the same order as the contribution to the imaginary part from these higher order processes which must be then taken into account if the available energy is larger than the threshold for sterile-like production.

The index of refraction, namely the real part of the self-energy is dominated by the one-loop result, which for low temperatures (large τ) is given by the low temperature limits (11.5.6,11.5.11). This observation is important: at $T \ll M_w$ on dimensional grounds the two loop processes yield real and imaginary parts of the self-energy $\propto G_F^2T^4q$ since this limit is well describe by the local Fermi theory, therefore compared to the low temperature limit of the *one loop* contribution (11.5.6,11.5.11) the two-loop contribution to the real part

is suppressed by a power of α_w. Therefore for $T \ll M_w$ the leading contribution to real part or index of refraction is given by the one loop results (11.5.6,11.5.11), whereas the imaginary part (damping rate) is determined by the two loop diagrams and are $\propto G_F^2 T^4 q$ the proportionality constant determined by the nature and number of degrees of freedom (leptons, quarks) that enter in the processes. Therefore, in principle a complete description of production and freeze out should include all possible processes at one and two loops in the self-energy. The real part is dominated by the one loop term, but the imaginary part will receive contributions from both one and two loops, the relevance of each will depend on the temperature regime. For the mixing angle, both the real and imaginary part (damping rate) are needed, however the imaginary part is of the same order than the real part only for the one-loop contribution, namely at temparatures of the order of $M_{w,z}$ (or larger), however, at much lower temperatures, the corrections to the mixing angle are dominated by one loop contribution to the real part given by (11.5.7,11.5.12) and the two loop contributions to both the real and imaginary part can be safely neglected in agreement with the results of ref.[415, 433, 346].

Therefore, the results obtained in the previous section provide a *lower bound* to the abundance of sterile-like neutrinos, as processes that are of higher order but dominate at lower temperatures increase the abundance.

11.8.3 Lifetime constraints:

Massive sterile-like neutrinos can decay in various leptonic channels[89, 399]. Consider the simpler case of one sterile-like ν_2 and one active-like ν_1 neutrino with $\theta \ll 1$, the charged current channel $\nu_2 \to e^+ e^- \nu_1$, is available for $M_2 \simeq M_s > 1\,\mathrm{MeV}$ and the "invisible" neutral current $\nu_2 \to 3\nu_1$ channel which is available for any M_s of cosmological relevance for WDM or CDM, and the radiative channel $\nu_2 \to \gamma\,\nu_1$ which is suppressed by one power of α_{em}. The decay widths for these channels have been obtained[89, 399, 452]

$$\Gamma(\nu_2 \to e^+ e^- \nu_1) \simeq 3.5 \times 10^{-5}\,\theta^2 \left(\frac{M_s}{MeV}\right)^5 K\left[\frac{m_e^2}{M_s^2}\right] \left(\frac{1}{s}\right) \tag{11.8.2}$$

where the function $K \to 0$ for $M_s \to 2m_e$ and $K \to 1$ for $M_s \gg m_e$[89]. For other leptonic channels similar expressions were obtained in ref.[452].

The decay rate into active-like neutrinos mediated by neutral currents (not GIM (Glashow Iliopoulos Maiani)) suppressed with sterile-like heavy neutrinos) is given by (see [89, 399])

$$\Gamma(\nu_2 \to 3\nu_1) \simeq 3.5 \times 10^{-5}\, \theta^2 \left(\frac{M_s}{MeV}\right)^5 \left(\frac{1}{s}\right) \tag{11.8.3}$$

and the radiative decay width[414, 399]

$$\Gamma(\nu_2 \to \gamma\nu_1) \simeq 10^{-7}\theta^2 \left(\frac{M_s}{\text{MeV}}\right)^5 \left(\frac{1}{s}\right). \tag{11.8.4}$$

Recent results for a lower bound on the lifetime of Dark Matter yields $t_b \simeq 160\,\text{Gyr}$[453], a similar bound but in terms of the fraction of cold dark matter is given in ref.[454]. Adding both leptonic channels assuming that $M_s \gtrsim$ MeV and taking both of the same order, the condition that the sterile species would be a suitable dark matter candidate implies that its lifetime is longer than or equal to this lower bound, namely $\Gamma_{tot}\, t_b \leq 1$, implying that

$$\theta^2 \left(\frac{M_s}{MeV}\right)^5 \lesssim 10^{-14}. \tag{11.8.5}$$

Combining this bound with the fractional abundance (11.7.27) we find that

$$\frac{\mathcal{F}_2 \left(\frac{M_s}{\text{MeV}}\right)^2}{\left[1 + \left(\frac{M_s}{8.35\,\text{MeV}}\right)^2\right]} \lesssim 10^{-6} \tag{11.8.6}$$

which could yield $\mathcal{F}_2 \simeq 1$ for $M_s \simeq$ few KeV, which, however would require a very large mixing angle $\theta \simeq 10^{-2}$ which is ruled out by cosmological X-ray bounds[393, 351, 352, 357]. Hence, we conclude that sterile-like neutrinos produced via vector boson decay cannot be suitable dark matter candidates.

However, if there is a hierarchy of sterile-like neutrinos heavy neutrinos with M_s few MeV and mixing angles $\theta^2 \gg 10^{-13}$, these may decay into lighter $\simeq$ KeV sterile-like states that could contribute to the dark matter abundance. This possibility of cascade decay merits further study and is clearly beyond the scope of this article.

11.8.4 Comparison to other results

The expressions for the quantum kinetic equation (11.3.18) and the effective mixing angle in the medium (11.2.43) that we obtained are *exact* to all orders in standard model couplings and to leading order in $\theta^2 \ll 1$. In ref.[455, 392] an expression for the effective mixing angle that includes both the real part (index of refraction) and imaginary part (damping rate) of the self-energy (from standard model interactions) has been proposed that seems to be valid for arbitrary vacuum mixing angle. Our result is only valid for $\theta \ll 1$ where we can extract unambiguously the mixing angle from the position of the complex poles in the propagator. It is not straightforward to define or extract a real mixing angle in the case of large θ, the subtleties are discussed in ref.[416]. The final form of the quantum kinetic equation (11.3.18) is similar to that used in [433, 69], although in these articles the mixing angle only includes the index of refraction (real part), valid in the temperature regime of interest in those articles. Crucially, our analysis shows the importance of the positive helicity component: at high temperature the in-medium suppression of the mixing angle for negative helicity is much larger than that for positive helicity since the latter interacts with the medium with a coupling that is helicity suppressed. However the contributions from the in-medium suppressed negative helicity and the helicity suppressed positive helicity may be competitive within a range of masses and temperatures. This is an aspect that has not been discussed previously. Finally, our results seem to be in broad agreement with those of ref.[390], although it is not clear in this ref. the role played by the positive helicity component, which as analyzed above, may play an important role in the production of *heavy sterile neutrinos*. Furthermore, this latter reference ultimately discards the one loop contribution as it focuses on the temperature regime $T \simeq 150\,\text{MeV}$ and $M_s \simeq$ KeV wherein the two loop contribution dominates.

11.8.5 Thermalization?

If the lifetime of the sterile-like neutrino is (much) shorter than the age of the Universe, it means that at some time in the past history of the Universe the *rate* $dn_s(t)/dt < 0$ since

if such species is present today its population is decaying in time. We have argued that the quantum kinetic equation (11.3.18) is *exact* to all orders in standard model couplings to leading order in $\theta^2 \ll 1$. In fact as per the discussion leading up to (11.3.18) the production term is completely determined by the damping rate and detailed balance, which if the sterile-like mass eigenstate is *not* relativistic entails that the correct form of the quantum kinetic equation is

$$\frac{dn_2^h(q;t)}{dt} = \Gamma_2^h(q)\Big[n_{LTE}(E) - n_2^h(q;t)\Big]\,, \tag{11.8.7}$$

where

$$n_{LTE}(E) = \frac{1}{e^{\frac{E}{T}}+1} \quad ; \quad \frac{E}{T} = \sqrt{y^2 + \frac{M_2^2}{T^2}} \tag{11.8.8}$$

and $\Gamma_2^h(q)$ are the damping rates in terms of the imaginary part of the self-energy (11.2.58). In Minkowski space time, however small the mixing angle (hence Γ_2) sterile-like mass eigenstates will *always* thermalize, the longer the thermalization time scale the smaller Γ_2. With cosmological expansion freeze out occurs when $\Gamma_2/H(t) \ll 1$ (for a more detailed discussion see ref.[447]). In the original form (gain-loss) of the kinetic equation (11.3.4) the gain term always involves the annihilation of one or several species (vector bosons, leptons,quarks) that by assumption are in (LTE) in the plasma, as the Universe expands and cools the abundances of these species diminishes and the gain contributions diminish accordingly. The loss terms that involve the annihilation of one or more species in LTE also diminish under cosmological expansion, however, if the sterile neutrino can decay into other species, this decay contribution only entails the *creation* of the decay products, and these contributions *do not* vanish as the temperature diminishes. Three processes that contribute to the loss term and survive in the low temperature limit are precisely the decay channels (11.8.2,11.8.3,11.8.4). Therefore if the sterile-like mass eigenstate decays with a lifetime smaller than the age of the Universe, these loss terms dominate the quantum kinetic equation at some late time and the rate becomes negative before today. Then the form (11.8.7) implies that at some time in the past the sterile-like neutrino has *thermalized*, since the production term (gain) dominates initially but the decay (loss) dominates at late times, the rate must have passed through zero in between, namely the distribution reached (LTE) and started to decay after this point. This discussion becomes relevant with the possibility that M_s few MeV and lifetime $\simeq 10^5\,(s)$

could provide a solution to the ^7Li problem as suggested in refs.[437, 441, 442], as discussed below.

11.8.6 Solution to the ^7Li problem?:

In ref.[442] the authors performed an exhaustive analysis of the parameter space within which the decay of sterile neutrinos of $M_s \simeq$ few MeV could yield a solution of the ^7Li problem as previously advocated in refs.[437, 441]. The analysis of ref.[442] included the most recent data on (CMB) anisotropies and concluded that a heavy sterile neutrino with $M_s \simeq 4.35\,\text{MeV}$ and lifetime $\Gamma^{-1} \simeq 1.8 \times 10^5\,(s)$ would provide a suitable solution. However, the parameter space also bounds the ratio[2] $\mathcal{N}_s/\mathcal{N}_\nu \simeq 10^{-4}$ and the mixing angle $\theta^2 \simeq 10^{-4}$. These values are in significant tension both with the results that we obtained above and the bounds of ref.[445]. In particular with M_s few MeV and such large mixing angle our result (11.7.25) indicates that $\mathcal{N}_s/\mathcal{N}_\nu \simeq 1$ suggesting full thermalization, furthermore, as is discussed above, our results provide a *lower bound* for the abundance of sterile-like heavy neutrinos. Both the region of abundance and mixing angles found in ref.[442] seem in strong tension with the bounds in[445], both caveats are recognized in[442] which suggests, as possible alternative, a low reheating temperature[436]. Of course our results rely on (LTE) at the electroweak scale, therefore they are not applicable to such scenario. Hence, although the production mechanism of sterile-like neutrinos studied in this article which is the leading order in standard model couplings and provides a lower bound to the abundance, offers a compelling mechanism for production of heavy sterile-like neutrinos with the possibility to solve the ^7Li problem, significant tension arises between the parameter range of the solution established in ref.[442], our result as a lower bound on the abundance and the cosmological bounds obtained in ref.[445]. The resolution of this tension merits a deeper study, well beyond the scope of this article.

[2]This reference actually bounds $\mathcal{N}_s/\mathcal{N}_{cmb}$ which differs by a factor 3/4.

11.8.7 WDM from cascade decay:

The analysis of the solution of the ^7Li problem suggested in in refs.[437, 441, 442] is a specific example of a cascade decay mechanism: heavy ($M_s \sim$ few MeV) sterile-like neutrinos produced at a (high) scale that eventually decay into several channels with the daughter particles influencing important physical processes during cosmological expansion. If there is a hierarchy of sterile-like massive neutrinos that include MeV and KeV scales, the heavier mass states may be produced at a high temperature, such as explored in this article, and the decay of this heavy state on a time scale $\simeq 10^5\, s$ to solve the ^7Li problem (if the caveats discussed above can be overcome), can also lead to the production of the lighter mass states that can be suitable WDM candidates. While this lifetime is interesting within the context of the ^7Li problem, a heavy neutrino with $M_s \simeq$ MeV and $\theta \simeq 10^{-7}$ would feature a lifetime $\simeq 10^{12}\,(s)$ therefore decaying into a WDM candidate just after matter radiation equality. This possibility emerges naturally by writing the weak interaction vertices in mass eigenstates, then the process $\nu_2 \to 3\nu_1$ yields a contribution $\nu_2 \to 2\nu_1\nu_m$ with ν_2, ν_m the heavier ($\simeq$ MeV) and lighter ($\simeq$ KeV) mass eigenstates respectively. The branching ratio for such process is $\propto \theta_m^2$ where θ_m^2 is the mixing angle of the active (flavor) neutrinos with the sterile-like lighter neutrino ν_m. This mechanism of production of WDM candidates is a tantalizing possibility that would be a natural scenario in extensions beyond the standard model that posit the existence of several sterile neutrinos merits further study clearly beyond the scope of this article.

11.9 SUMMARY OF RESULTS, CONCLUSIONS, FURTHER QUESTIONS

Our goals in this article are two-fold: i) to obtain the general form of the quantum kinetic equations and effective mixing angles in the medium to describe production and freeze-out of sterile-like (mass eigenstates) neutrinos in a broad range of temperature and under a minimal set of assumptions, ii) to apply these to study the production to leading order in standard model couplings from vector boson decay at $T \simeq M_w$.

We obtained the effective mixing angles in the medium directly from the equations of motion in the case of mixing of one sterile with one active neutrino via a see-saw mass matrix with standard model interactions for the active (flavor) neutrino valid when the vacuum mixing angle $\theta \ll 1$ but to *all orders in standard model couplings*. Assuming that all standard model degrees of freedom are in (LTE) in the relevant temperature range we obtained the quantum kinetic equation that describes the production, evolution and freeze-out of sterile-like mass eigenstates. The mixing angles in the medium and the production rate are determined by the real and imaginary parts of the self-energy on the mass shell of the sterile-like mass eigenstate, and depend on helicity. The full quantum kinetic equation to leading order in $\theta \ll 1$ is

$$\frac{dn_2^h(q;t)}{dt} = \Gamma_2^h(q)\Big[n_{LTE}(q) - n_2^h(q;t)\Big] \,,$$

where $h = \pm$ correspond to helicity states and $\Gamma_2^{\mp}(q)$ are given by (11.2.58) with (11.2.54, 11.2.54), and n_{LTE} is the Fermi-Dirac distribution function in (LTE). The full expression for the mixing angles in the medium, valid to all orders in standard model couplings and to leading order in $\theta \ll 1$ is given in the relativistic limit by

$$\theta^h_{eff}(q) = \frac{\theta}{\left[\left(1+\frac{\Delta^h(q)}{\xi}\right)^2 + \left(\frac{\gamma^h(q)}{\xi}\right)^2\right]^{1/2}} \,,$$

where Δ, γ, ξ are given by (11.2.54-11.2.57,11.2.40) respectively in terms of the real (Δ) and imaginary (γ) part of the *active neutrino self-energy on the mass shell of the sterile-like eigenstate*.

We implemented the quantum kinetic equation to obtain the production of sterile-like neutrinos from vector boson decay at $T \simeq M_w$ including cosmological expansion. For negative helicity neutrinos (and positive helicity anti-neutrinos) the effective mixing angle is strongly suppressed by the medium, however for positive helicity neutrinos (and negative helicity anti-neutrinos) the medium corrections are negligible because the interaction with the medium is helicity suppressed. We find that there is a region of masses for which the production of both species is comparable.

The mixing angle for negative helicity neutrinos features an MSW resonance *in absence of lepton asymmetry*, which, however, is screened by the imaginary part of the self-energy. Negative helicity neutrinos freeze-out at $T_f^- \simeq 5\,\mathrm{GeV}$ with a broader distribution as a consequence of a competition between a diminishing damping rate γ and an *increasing* effective mixing angle as temperature diminishes. Positive helicity neutrinos freeze-out temperature is $T_f^+ \simeq 8\,\mathrm{GeV}$ with a distribution that peaks at much smaller momenta, describing a colder species. Accounting for both channels we find that the distribution function of sterile-like neutrinos of mass $M_2 \simeq M_s$ is given by

$$n_2(y) = 3.6 \left(\frac{\theta^2}{10^{-4}}\right)\left(\frac{M_s}{\mathrm{MeV}}\right)^2 f(M_s, y)\,,$$

where $y = q/T$ and $y^2 f(M_s, y)$ is strongly non-thermal and is displayed in fig.(83) revealing the competition between the colder (positive helicity) and hotter (negative helicity) components. The total abundance normalized to that of one relativistic degree of freedom in thermal equilibrium ($\mathcal{N}_\nu$) is

$$\frac{\mathcal{N}_2}{\mathcal{N}_\nu} \simeq 2 \left(\frac{\theta^2}{10^{-4}}\right) \left(\frac{M_s}{\mathrm{MeV}}\right)^2 \left[1 + \left(\frac{M_s}{8.35\,\mathrm{MeV}}\right)^2\right]. \tag{11.9.1}$$

The first term in the bracket is the contribution from the positive helicity states and the second from the negative helicity, both become comparable for $M_s \simeq 8.35\,\mathrm{MeV}$. We argue that this expression is a *lower bound* on the abundance of sterile-like neutrinos.

The fractional abundance of dark matter contributed by both helicity components is given by (11.7.27). Constraints from X-ray data on masses and mixing angles suggest that sterile-like neutrinos produced by vector boson decay *cannot* yield a substantial $M_s \simeq$ KeV warm dark matter component. However, this production mechanism yield a substantial abundance of $M_s \simeq$ MeV *heavy* sterile-like neutrinos with $\theta^2 < 10^{-4}$ consistent with accelerator constraints. Therefore this production mechanism may yield the heavy neutrinos recently invoked to solve the ^7Li problem[437, 441, 442]. However, the parameter range determined in [442] also bounds $\mathcal{N}/\mathcal{N}_\nu \simeq 10^{-4}$ which is incompatible with the result (11.9.1) for the range of mass and mixing angles reported in this reference, and is also in conflict with recent bounds reported in[445]. The possibility that heavy $\simeq$ MeV sterile-like neutrinos

decaying after BBN injecting energy in the medium providing a solution of the ^7Li problem as suggested also in refs.[437, 441, 442] merits a deeper study both of the production mechanism as well as the cosmological impact of this heavy neutrino species.

Further questions: We have suggested several other processes that contribute to the production throughout the thermal history of the Universe, while these are higher order (two loops) processes they may comparable to the leading order processes or even dominate at temperature $T \ll M_w$. However the medium corrections to the mixing angles are completely determined by the one loop contribution and the effective mixing angle is given by eqn. (11.6.2). Further study of these processes is clearly warranted, they can be competitive near the freeze-out temperature of the leading one-loop contribution for heavy sterile neutrinos and, crucially contribute to the production of lighter mass eigenstates. We have also argued that sterile neutrinos with lifetimes shorter than the age of the Universe that are decaying today must necessarily have thermalized at some time in the past. Studying this thermalization process is of fundamental interest since most of the calculations of production neglect this possibility by neglecting the loss term in the kinetic equation, yet thermalization of heavy sterile neutrinos may have important cosmological consequence for the expansion history. We have also suggested that several compelling extensions beyond the standard model posit a hierarchy of sterile neutrino masses, and this opens the possibility that heavier sterile-states may be produced at high temperature, as analyzed here, and decay well after BBN or near the time of matter-radiation equality into lighter sterile states that may be suitable WDM candidates. This mechanism of cascade decay, which is fundamentally similar to that advocated for the solution of the ^7Li problem as an energy injection mechanism, is worthy of study.

11.10 APPENDICES

11.10.1 Spectral density

The spectral densities are obtained for a generic vector boson mass M, with a straightforward application for either charged or neutral current cases.

We need to identify the regions in which the product of delta functions in (11.4.23) with (11.4.23) have support in the interval $W^- \leq W \leq W^+$.

Using the identities:

$$n_f(-k) = 1 - n_f(k) \quad ; \quad N_b(-p^0) = -(1 + N_b(p^0)) \tag{11.10.1}$$

the product

$$\Big[\delta(k_0-k)-\delta(k_0+k)\Big]\,\Big[\delta(q_0-k^0-W_p)-\delta(q_0-k^0+W_p)\Big]\,\Big[1-n_f(k_0)+N_b(p_0)\Big] \quad ; \quad p^0 = q^0-k^0 \tag{11.10.2}$$

is gathered into four different terms, keeping the finite temperature contributions only these are:

$$\mathbf{1}): \quad \delta(k^0-k)\delta(q^0-k-W)\Big[-n_f(k)+N_b(q^0-k)\Big] \tag{11.10.3}$$

this has support for $W^- \leq q^0 - k \leq W^+$ as a function of k this constraint implies

$$k^- \leq k \leq k^+ \quad ; \quad k^\pm = \frac{(q^0)^2 - q^2 - M^2}{2(q^0 \mp q)} \quad \text{for} \quad q^0 \geq \sqrt{q^2+M^2} \tag{11.10.4}$$

$$\mathbf{2}): \quad \delta(k^0+k)\delta(q^0+k+W)\Big[1-n_f(-k)+N_b(q^0+k)\Big]\,, \tag{11.10.5}$$

For this term it must be that $q^0 + k < 0$ since $W^- \leq W \leq W^+$ and because $k > 0$, it follows that $q^0 < 0$. Therefore using the identity (11.10.1) and keeping solely the finite temperature contributions this term yields

$$\mathbf{2}): \quad -\delta(k^0+k)\delta(|q^0|-k-W)\Big[-n_f(k)+N_b(|q^0|-k)\Big] \quad ; \quad |q^0|-k>0\,, \tag{11.10.6}$$

the region of support is $q^0 < 0$ and for k is

$$k^- \leq k \leq k^+ \;\; ; \;\; k^\pm = \frac{(q^0)^2 - q^2 - M^2}{2(|q^0| \mp q)} \;\; \text{for} \;\; |q^0| \geq \sqrt{q^2 + M^2} \tag{11.10.7}$$

$$\mathbf{3}) : -\delta(k^0 - k)\delta(q^0 - k + W)\Big[1 - n_f(k) + N_b(q^0 - k)\Big] \tag{11.10.8}$$

this term has support for $k - q^0 > 0$, using (11.10.1)

$$\mathbf{3}) : \delta(k^0 - k)\delta(k - q^0 - W)\Big[n_f(k) + N_b(k - q^0)\Big]\,, \tag{11.10.9}$$

the regions of support are:

$$i :) \qquad 0 \leq q^0 \leq q \;;\; k^- \leq k < \infty \;\; ; \;\; k^- = \frac{q^2 + M^2 - (q^0)^2}{2(q - q^0)} \tag{11.10.10}$$

$$ii :) \qquad 0 > q^0 > -q \;\; ; \;\; k^+ = \infty \;\; ; \;\; k^- = \frac{q^2 + M^2 - (q^0)^2}{2(q - q^0)} \tag{11.10.11}$$

$$iii :) \qquad q^0 < 0 \;;\; q \leq |q^0| \leq \sqrt{q^2 + M^2} \;\; ; \;\; k^- \leq k \leq k^+ \tag{11.10.12}$$
$$k^\pm = \frac{q^2 + M^2 - (q^0)^2}{2(|q^0| \mp q)}$$

$$\mathbf{4}) : -\delta(k^0 + k)\delta(q^0 + k - W)\Big[n_f(k) + N_b(q^0 + k)\Big]\,, \tag{11.10.13}$$

the regions of support are:

$$i :) \; 0 \leq q^0 \leq q : \;\; ; \;\; k^- \leq k < \infty \;\; ; \;\; k^- = \frac{q^2 + M^2 - (q^0)^2}{2(q + q^0)} \tag{11.10.14}$$

$$ii :) \; q \leq q^0 \leq \sqrt{q^2 + M^2} : \;\; ; \;\; k^- \leq k \leq k^+ \;\; ; \;\; k^\pm = \frac{q^2 + M^2 - (q^0)^2}{2(q \mp q^0)} \tag{11.10.15}$$

$$iii :) \; -q \leq q^0 \leq 0 : \;\; ; \;\; k^- \leq k < \infty \;\; ; \;\; k^- = \frac{q^2 + M^2 - (q^0)^2}{2(q + q^0)} \tag{11.10.16}$$

The integrals over k^0 can be done straightforwardly, the contributions from $L[Q,k]\cdot P[p]$ (see (11.4.12,11.4.15) yield the following terms for negative and positive helicity respectively:

k^0 integrals: Negative helicity

- for **1** : $-\frac{M^2}{q}\Big[F_1(q_0,q)+k\,F_2(q_0,q)\Big]$
- for **2** : $-\frac{M^2}{q}\Big[F_1(q_0,q)-k\,F_2(q_0,q)\Big]$
- for **3** : $-\frac{M^2}{q}\Big[F_1(q_0,q)+k\,F_2(q_0,q)\Big]$
- for **4** : $-\frac{M^2}{q}\Big[F_1(q_0,q)-k\,F_2(q_0,q)\Big]$

k^0 integrals: Positive helicity

- for **1** : $\frac{M^2}{q}\Big[G_1(q_0,q)+k\,G_2(q_0,q)\Big]$
- for **2** : $\frac{M^2}{q}\Big[G_1(q_0,q)-k\,G_2(q_0,q)\Big]$
- for **3** : $\frac{M^2}{q}\Big[G_1(q_0,q)+k\,G_2(q_0,q)\Big]$
- for **4** : $\frac{M^2}{q}\Big[G_1(q_0,q)-k\,G_2(q_0,q)\Big]$

k Integrals

The next step is to calculate the $k-$ integrals, this is facilitated by the following identities:

$$n_f(k) = -T\frac{d}{dk}\ln[1+e^{-k/T}] \tag{11.10.17}$$

$$N_b(q^0-k) = -T\frac{d}{dk}\ln[1-e^{k/T}e^{-q^0/T}] \tag{11.10.18}$$

and a similar identity for $N_b(k-q^0); N_b(q^0+k)$. With these identities we find

$$\int_{k^-}^{k^+} n_f(k)\;dk = -T\;\ln\left[\frac{1+e^{-k^+/T}}{1+e^{-k^-/T}}\right] \tag{11.10.19}$$

$$\int_{k^-}^{k^+} N_b(q^0-k)\;dk = -T\;\ln\left[\frac{1-e^{k^+/T}e^{-q^0/T}}{1-e^{k^-/T}e^{-q^0/T}}\right] \tag{11.10.20}$$

$$\int_{k^-}^{k^+} N_b(k-q^0)\;dk = T\;\ln\left[\frac{1-e^{-k^+/T}e^{q^0/T}}{1-e^{-k^-/T}e^{q^0/T}}\right] \tag{11.10.21}$$

$$\begin{aligned}\int_{k^-}^{k^+} k\;n_f(k)\;dk &= -T^2\left\{k^+\ln\left[1+e^{-k^+/T}\right]-k^-\ln\left[1+e^{-k^-/T}\right]\right.\\ &-\left.\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\Big[e^{-nk^+/T}-e^{-nk^-/T}\Big]\right\}\end{aligned} \tag{11.10.22}$$

$$\int_{k^-}^{k^+} k\ N_b(q^0-k)\ dk \;=\; -T^2\left\{k^+\ln\left[1-e^{k^+/T}e^{-q^0/T}\right]-k^-\ln\left[1-e^{k^-/T}e^{-q^0/T}\right]\right.$$
$$\left.+\ \sum_{n=1}^{\infty}\frac{e^{-nq^0/T}}{n^2}\left[e^{nk^+/T}-e^{nk^-/T}\right]\right\} \tag{11.10.23}$$

$$\int_{k^-}^{k^+} k\ N_b(k-q^0)\ dk \;=\; -T^2\left\{-k^+\ln\left[1-e^{-k^+/T}e^{q^0/T}\right]+k^-\ln\left[1-e^{-k^-/T}e^{q^0/T}\right]\right.$$
$$\left.+\ \sum_{n=1}^{\infty}\frac{e^{nq^0/T}}{n^2}\left[e^{-nk^+/T}-e^{-nk^-/T}\right]\right\} \tag{11.10.24}$$

In the integrals (11.10.22-11.10.24) we have used the identities (11.10.17,11.10.18), integrated by parts, expanded the logarithms in power series and integrated term by term. The infinite sums can be expressed in terms of di-logarithmic (Spence's) functions but such form is not particularly useful.

Numerical Implementation for the real part:

The numerical implementation for the real parts from the dispersive form (11.2.13) is best achieved in a "modular form" which is facilitated by introducing

$$\int_{k^-}^{k^+}\left[N_b(q^0-k)-n_f(k)\right]dk \;\equiv\; -T\,D1(q^0,q) \tag{11.10.25}$$
$$\int_{k^-}^{k^+}\left[N_b(q^0-k)-n_f(k)\right]k\,dk \;\equiv\; -T^2\,Dk1(q^0,q) \tag{11.10.26}$$
$$\int_{k^-}^{k^+}\left[N_b(k-q^0)+n_f(k)\right]dk \;\equiv\; -T\,D2(q^0,q) \tag{11.10.27}$$
$$\int_{k^-}^{k^+}\left[N_b(k-q^0)+n_f(k)\right]k\,dk \;\equiv\; -T^2\,Dk2(q^0,q)\,, \tag{11.10.28}$$

where the respective integrals are given above. In terms of these quantities and $F_{1,2}(q^0,q)$ defined by eqns. (11.4.13,11.4.14) (and $G_{1,2}$ defined by (11.4.16,11.4.17) for positive helicity) we find for negative helicity the following contributions to the imaginary parts in the different regions $1-4$ of q^0 defined by the support of the corresponding delta functions described above:

$$\mathrm{Im}\Sigma^{(1)}=\frac{g^2M^2T}{16\pi\,q^2}\left[F_1(q^0,q)\,D1(q^0,q)+T\,F_2(q^0,q)\,Dk1(q^0,q)\right]\quad;\quad q^0\geq\sqrt{q^2+M^2} \tag{11.10.29}$$

$$\begin{aligned}
\mathrm{Im}\Sigma^{(2)} &= -\frac{g^2 M^2 T}{16\pi\, q^2}\left[F_1(q^0,q)\, D1(|q^0|,q) - T\, F_2(q^0,q)\, Dk1(|q^0|,q)\right] \\
q^0 < 0 \quad &, \quad |q^0| \geq \sqrt{q^2+M^2}
\end{aligned} \tag{11.10.30}$$

$$\begin{aligned}
\mathrm{Im}\Sigma^{(3)} &= \frac{g^2 M^2 T}{16\pi\, q^2}\Big\{\Big[F_1(q^0,q)\, D2(q^0,q) + T\, F_2(q^0,q)\, Dk2(q^0,q)\Big](0<q^0<q) \\
&+ \Big[F_1(q^0,q)\, D2(q^0,q) + T\, F_2(q^0,q)\, Dk2(q^0,q)\Big](0>q^0>-q) \\
&+ \Big[F_1(q^0,q)\, D2(q^0,q) + T\, F_2(q^0,q)\, Dk2(q^0,q)\Big](-q>q^0>-\sqrt{k^2+M^2})\Big\}
\end{aligned} \tag{11.10.31}$$

in the first two terms $k^+ = \infty \to D2(k^+=\infty) = Dk2(k^+=\infty) = 0$ and only the lower limit with k^- corresponding to the case (3) above contributes, the two limits $k^\pm$ contribute to the last term.

$$\begin{aligned}
\mathrm{Im}\Sigma^{(4)} &= -\frac{g^2 M^2 T}{16\pi\, q^2}\Big\{\Big[F_1(q^0,q)\, D2(-q^0,q) - T\, F_2(q^0,q)\, Dk2(-q^0,q)\Big](0<q^0<q) \\
&+ \Big[F_1(q^0,q)\, D2(-q^0,q) - T\, F_2(q^0,q)\, Dk2(-q^0,q)\Big](q<q^0<\sqrt{k^2+M^2}) \\
&+ \Big[F_1(q^0,q)\, D2(-q^0,q) - T\, F_2(q^0,q)\, Dk2(-q^0,q)\Big](0>q^0>-q)\Big\}\,.
\end{aligned} \tag{11.10.32}$$

Now the real part of the self energy is calculated with the dispersive form (11.2.13) with

$$\mathrm{Im}\Sigma(q^0,q) = \mathrm{Im}\Sigma^{(1)}(q^0,q) + \mathrm{Im}\Sigma^{(2)}(q^0,q) + \mathrm{Im}\Sigma^{(3)}(q^0,q) + \mathrm{Im}\Sigma^{(4)}(q^0,q)\,. \tag{11.10.33}$$

In each region in q^0 the values of $k^\pm$ are given by the different cases analyzed above. The principal part is obtained by excising an interval of width 2ϵ around $q^0 = q$ with $\epsilon \ll 1$.

For positive helicity the same analysis above holds with the following modifications:

- **i:)** $M^2 \to -M^2$ *only in the pre-factor*
- **ii:)** $F_{1,2}(q^0,q) \to G_{1,2}(q^0,q)$ where $F_{1,2}, G_{1,2}$ are given by eqns. (11.4.13,11.4.14) and (11.4.16,11.4.17) respectively. Note that $F_{1,2}$ and $G_{1,2}$ obey the relation (11.4.18).

Finally, we introduce the dimensionless variables:

$$z = q^0/T \;\; ; \;\; y = q/T \;\; ; \;\; \tau = M_w/T \tag{11.10.34}$$

where as discussed in the text we use M_w as the baseline scale. The integrals over q^0 are then rendered dimensionless in terms of these variables. The dispersive integrals over the dimensionless variable z are carried out numerically and the final results for the real part of the self energy are generically of the form

$$\mathrm{Re}\Sigma^{\pm}(q) = \frac{g^2 M^2 T}{16\pi^2\, q^2}\, \mathcal{K}^{\pm}[\tau, y]\,, \tag{11.10.35}$$

where $\mathcal{K}^{\pm}[\tau, y]$ are dimensionless functions of τ, y that are obtained numerically with the procedure detailed above. For charged currents $g^2 = g_w^2/2, M = M_w$, for neutral currents $g^2 = g_w^2/(2c)^2, M = M_w/c, c = \cos(\theta_w) \simeq 0.88$ and for neutral currents $\tau \to \tau/c$ in the argument of $\mathcal{K}[\tau, y]$.

Low Temperature limit.

In the low temperature limit $M_w \gg T$ and keeping only the finite temperature contributions and using the identities (11.10.1) we can neglect N_b. In (11.4.9) the product of delta functions and distribution functions becomes

$$n_f(k)\Big\{\delta(k^0-k)\Big[\delta(q^0-k+W_p)-\delta(q^0-k-W_p)\Big]+\delta(k^0+k)\Big[\delta(q^0+k+W_p)-\delta(q^0+k-W_p)\Big]\Big\}\,. \tag{11.10.36}$$

Now it is more convenient to integrate over q^0 and k^0 in the dispersive integral (11.2.13), leaving only the integrals in k, with

$$d^3k = (2\pi)k^2 dk d(\cos(\varphi)) \tag{11.10.37}$$

with φ the angle between $\vec{q}$ and $\vec{k}$.

For negative helicity we find

$$L_{\mu\nu}[Q^+, k]P^{\mu\nu}[p] = k(1-\cos(\varphi)) + \frac{2k}{M^2}(q^0 - q\cos(\varphi))\Big(q^0 - k - q + k\cos(\varphi)\Big) \;\; ; \;\; \text{for} \;\; k^0 = k \tag{11.10.38}$$

$$L_{\mu\nu}[Q^+,k]P^{\mu\nu}[p] = -k(1+\cos(\varphi))-\frac{2k}{M^2}(q^0+q\cos(\varphi))\Big(q^0+k-q+k\cos(\varphi)\Big) \;\; ; \;\; \text{for} \;\; k^0=-k\,. \tag{11.10.39}$$

Integrating over q^0 implementing the delta functions and expanding the numerator and denominator in powers of $k/M, q/M$, integrating over $\cos(\varphi)$, keeping only the leading order terms (proportional to $1/M^4$) and using

$$\int_0^\infty k^3\, n_f(k)\, dk = \frac{7\,\pi^4\,T^4}{120} \tag{11.10.40}$$

we find for negative helicity

$$\mathrm{Re}\Sigma^-(q) = \frac{14\pi^2}{90}\, g^2 \left(\frac{T}{M}\right)^4 q \tag{11.10.41}$$

For positive helicity we follow the same steps, with

$$L_{\mu\nu}[Q^-,k]P^{\mu\nu}[p] = k(1+\cos(\varphi))+\frac{2k}{M^2}(q^0-q\cos(\varphi))\Big(q^0-k+q-k\cos(\varphi)\Big) \;\; ; \;\; \text{for} \;\; k^0=k \tag{11.10.42}$$

$$L_{\mu\nu}[Q^-,k]P^{\mu\nu}[p] = -k(1-\cos(\varphi))-\frac{2k}{M^2}(q^0+q\cos(\varphi))\Big(q^0+k+q-k\cos(\varphi)\Big) \;\; ; \;\; \text{for} \;\; k^0=-k\,. \tag{11.10.43}$$

Following the same steps as for negative helicity, we find

$$\mathrm{Re}\Sigma^+(q) = \frac{14\pi^2}{180}\, g^2 \left(\frac{T}{M}\right)^4 q\,. \tag{11.10.44}$$

12.0 SUMMARY OF THE MAIN RESULTS

Here we provide a brief summary of the main results of this thesis.

Searching for steriles from π/K decay

The production of heavy sterile neutrinos from π^-, K^- decay at rest yields charged leptons with *negative* helicity (*positive* for π^+, K^+). We obtain the branching ratio for this process and argue that a Stern-Gerlach filter with a magnetic field gradient leads to spatially separated domains of both helicity components with abundances determined by the branching ratio. Complemented with a search of the monochromatic peak, this setup can yield both the mass and mixing angles for sterile neutrinos with masses in the range $3\,\text{MeV} \lesssim m_s \lesssim 414\,\text{MeV}$ in next generation high intensity experiments. We also study oscillations of light Dirac and Majorana sterile neutrinos with $m_s \simeq$ eV produced in meson decays including decoherence aspects arising from lifetime effects of the decaying mesons and the stopping distance of the charged lepton in short baseline experiments. We obtain the transition probability from production to detection via charged current interactions including these decoherence effects for $3+1$ and $3+2$ scenarios, also studying $|\Delta L| = 2$ transitions from $\overline{\nu} \leftrightarrow \nu$ oscillations for Majorana neutrinos and the impact of these effects on the determination of CP-violating amplitudes. We argue that decoherence effects are important in current short baseline accelerator experiments, leading to an *underestimate* of masses, mixing and CP-violating angles. At MiniBooNE/SciBooNE we estimate that these effects lead to an $\sim 15\%$ underestimate for sterile neutrino masses $m_s \gtrsim 3\,\text{eV}$. We argue that reactor and current short baseline accelerator experiments are fundamentally different and suggest that in future high intensity experiments with neutrinos produced from π, K

decay at rest, stopping the charged leptons on distances much smaller than the decay length of the parent meson suppresses considerably these decoherence effects.

Charged lepton mixing via heavy sterile neutrinos

Pseudoscalar meson decay leads to an entangled state of charged leptons (μ, e) and massive neutrinos. Tracing out the neutrino degrees of freedom leads to a reduced density matrix for the charged leptons whose off-diagonal elements reveal *charged lepton oscillations*. Although these decohere on unobservably small time scales $\lesssim 10^{-23}s$ they indicate charged lepton *mixing* as a result of common intermediate states. The charged lepton self energy up to one loop features flavor off-diagonal terms responsible for charged lepton mixing: a dominant "short distance" contribution with W bosons and massive neutrinos in the intermediate state, and a subdominant "large distance" contribution with pseudoscalar mesons and massive neutrinos in the intermediate state. Mixing angle(s) are GIM suppressed, and are *momentum and chirality dependent*. The difference of negative chirality mixing angles near the muon and electron mass shells is $\theta_L(M_\mu^2) - \theta_L(M_e^2) \propto G_F \sum U_{\mu j} m_j^2 U_{je}^*$ with m_j the mass of the neutrino in the intermediate state. Recent results from TRIUMF, suggest an upper bound $\theta_L(p^2 \simeq M_\mu^2) - \theta_L(p^2 \simeq M_e^2) < 10^{-14} \left(M_S/100\, MeV \right)^2$ for one generation of a heavy sterile neutrino with mass M_S. We obtain the wavefunctions for the propagating modes, and discuss the relation between the lepton flavor violating process $\mu \to e\gamma$ and charged lepton mixing, highlighting that a measurement of such process implies a mixed propagator μ, e. Furthermore writing flavor diagonal vertices in terms of mass eigenstates yields novel interactions suggesting further contributions to lepton flavor violating process as a consequence of momentum and chirality dependent mixing angles.

Entanglement entropy in particle decay

The decay of a parent particle into two or more daughter particles results in an entangled quantum state as a consequence of conservation laws in the decay process. Recent experiments at Belle and BaBar take advantage of quantum entanglement and the correlations in the time evolution of the product particles to study CP and T violations. If one (or more)

of the product particles are not observed, their degrees of freedom are traced out of the pure state density matrix resulting from the decay, leading to a mixed state density matrix and an entanglement entropy. This entropy is a measure of the loss of information present in the original quantum correlations of the entangled state. We use the Wigner-Weisskopf method to construct an approximation to this state that evolves in time in a *manifestly unitary* way. We then obtain the entanglement entropy from the reduced density matrix of one of the daughter particles obtained by tracing out the unobserved states, and follow its time evolution. We find that it grows over a time scale determined by the lifetime of the parent particle to a maximum, which when the width of the parent particle is narrow, describes the phase space distribution of maximally entangled Bell-like states. The method is generalized to the case in which the parent particle is described by a wave packet localized in space. Possible experimental avenues to measure the entanglement entropy in the decay of mesons at rest are discussed.

Superhorizon entanglement entropy from particle decay in inflation

In inflationary cosmology all particle states decay as a consequence of the lack of kinematic thresholds. The decay of an initial single particle state yields an *entangled quantum state of the product particles*. We generalize and extend a manifestly unitary field theoretical method to obtain the time evolution of the quantum state. We consider the decay of a light scalar field with mass $M \ll H$ with a cubic coupling in de Sitter space-time. Radiative corrections feature an infrared enhancement manifest as poles in $\Delta = M^2/3H^2$ and we obtain the quantum state in an expansion in Δ. To leading order the pure state density matrix describing the decay of a particle with sub-horizon wavevector is dominated by the emission of superhorizon quanta, describing *entanglement between superhorizon and subhorizon fluctuations and correlations across the horizon*. Tracing over the superhorizon degrees of freedom yields a mixed state density matrix from which we obtain the entanglement entropy. Asymptotically this entropy grows with the *physical* volume as a consequence of more modes of the decay products crossing the Hubble radius. A generalization to localized wave packets is provided. The cascade decay of single particle states into many particle states is

discussed. We conjecture on *possible* impact of these results on non-gaussianity and on the "low multipole anomalies" of the CMB.

Time evolution of cascade decay

We study non-perturbatively the time evolution of cascade decay for generic fields $\pi \to \phi_1\phi_2 \to \phi_2\chi_1\chi_2$ and obtain the time dependence of amplitudes and populations for the resonant and final states. We analyze in detail the different time scales and the manifestation of unitary time evolution in the dynamics of production and decay of resonant intermediate and final states. The probability of occupation (population) "flows" as a function of time from the initial to the final states. When the decay width of the parent particle Γ_π is much larger than that of the intermediate resonant state Γ_{ϕ_1} there is a "bottleneck" in the flow, the population of resonant states builds up to a maximum at $t^* = \ln[\Gamma_\pi/\Gamma_{\phi_1}]/(\Gamma_\pi - \Gamma_{\phi_1})$ nearly saturating unitarity and decays to the final state on the longer time scale $1/\Gamma_{\phi_1}$. As a consequence of the wide separation of time scales in this case the cascade decay can be interpreted as evolving sequentially $\pi \to \phi_1\phi_2;\ \phi_1\phi_2 \to \phi_2\chi_1\chi_2$. In the opposite limit the population of resonances (ϕ_1) does not build up substantially and the cascade decay proceeds almost directly from the initial parent to the final state without resulting in a large amplitude of the resonant state. An alternative but equivalent non-perturbative method useful in cosmology is presented. Possible phenomenological implications for heavy sterile neutrinos as resonant states and consequences of quantum entanglement and correlations in the final state are discussed.

Pre-slow roll initial conditions: infrared aspects and large scale power suppression during inflation

If the large scale anomalies in the temperature power spectrum of the cosmic microwave background are of primordial origin, they may herald modifications to the slow roll inflationary paradigm on the largest scales. We study the possibility that the origin of the large scale power suppression is a modification of initial conditions during slow roll as a result of a pre-slow roll phase during which the inflaton evolves rapidly. This stage is manifest in

a potential in the equations for the Gaussian fluctuations during slow roll and modify the power spectra of scalar and tensor perturbations via an initial condition transfer function $\mathcal{T}(k)$. We provide a general analytical study of its large and small scale properties and analyze the impact of these initial conditions on the infrared aspects of typical scalar field theories. The infrared behavior of massless minimally coupled scalar field theories leads to the dynamical generation of mass and anomalous dimensions, both depend non-analytically on $\mathcal{T}(0)$. During inflation all quanta decay into many quanta even of the same field because of the lack of kinematic thresholds. The decay leads to a quantum entangled state of sub and superhorizon quanta with correlations across the horizon. We find the modifications of the decay width and the entanglement entropy from the initial conditions. In all cases, initial conditions from a "fast-roll" stage that lead to a suppression in the scalar power spectrum at large scales also result in a suppression of the dynamically generated masses, anomalous dimensions and decay widths.

Tensor to scalar ratio and large scale power suppression from pre-slow roll initial conditions

We study the corrections to the power spectra of curvature and tensor perturbations and the tensor-to-scalar ratio r in single field slow roll inflation with standard kinetic term due to initial conditions imprinted by a "fast-roll" stage prior to slow roll. For a wide range of initial inflaton kinetic energy, this stage lasts only a few e-folds and merges smoothly with slow-roll thereby leading to non-Bunch-Davies initial conditions for modes that exit the Hubble radius during slow roll. We describe a program that yields the dynamics in the fast-roll stage while matching to the slow roll stage in a manner that is independent of the inflationary potentials. Corrections to the power spectra are encoded in a "transfer function" for initial conditions $\mathcal{T}_\alpha(k)$, $\mathcal{P}_\alpha(k) = P_\alpha^{BD}(k)\mathcal{T}_\alpha(k)$, implying a modification of the "consistency condition" for the tensor to scalar ratio at a pivot scale k_0: $r(k_0) = -8n_T(k_0)\left[\mathcal{T}_T(k_0)/\mathcal{T}_\mathcal{R}(k_0)\right]$. We obtain $\mathcal{T}_\alpha(k)$ to leading order in a Born approximation valid for modes of observational relevance today. A fit yields $\mathcal{T}_\alpha(k) = 1 + A_\alpha k^{-p}\cos[2\pi\omega k/H_{sr} + \varphi_\alpha]$, with $1.5 \lesssim p \lesssim 2$, $\omega \simeq 1$ and H_{sr} the Hubble scale during slow roll inflation, where curvature and tensor perturbations feature the *same* p, ω for a wide range of initial conditions. These corrections lead to both a

suppression of the quadrupole and oscillatory features in both $P_R(k)$ and $r(k_0)$ with a period of the order of the Hubble scale during slow roll inflation. The results are quite general and independent of the specific inflationary potentials, depending solely on the ratio of kinetic to potential energy κ and the slow roll parameters ϵ_V, η_V to leading order in slow roll. For a wide range of κ and the values of $\epsilon_V; \eta_V$ corresponding to the upper bounds from Planck, we find that the low quadrupole is consistent with the results from Planck, and the oscillations in $r(k_0)$ as a function of k_0 *could* be observable if the modes corresponding to the quadrupole and the pivot scale crossed the Hubble radius very few $(2-3)$ e-folds after the onset of slow roll. We comment on possible impact on the recent BICEP2 results.

Cosmological Implications of Light Sterile Neutrinos produced after the QCD Phase Transition

We study the production of sterile neutrinos in the early universe from $\pi \to l\nu_s$ shortly after the QCD phase transition in the absence of a lepton asymmetry while including finite temperature corrections to the π mass and decay constant f_π. Sterile neutrinos with masses $\lesssim 1MeV$ produced via this mechanism freeze-out at $T_f \simeq 10MeV$ with a distribution function that is highly non-thermal and features a sharp enhancement at low momentum thereby making this species *cold* even for very light masses. Dark matter abundance constraints from the CMB and phase space density constraints from the most dark matter dominated dwarf spheroidal galaxies provide upper and lower bounds respectively on combinations of mass and mixing angles. For $\pi \to \mu\nu_s$, the bounds lead to a narrow region of compatibility with the latest results from the 3.55KeV line. The non-thermal distribution function leads to free-streaming lengths (today) in the range of $\sim$ few kpc consistent with the observation of cores in dwarf galaxies. For sterile neutrinos with mass $\lesssim 1eV$ that are produced by this reaction, the most recent accelerator and astrophysical bounds on U_{ls} combined with the non-thermal distribution function suggests a substantial contribution from these sterile neutrinos to N_{eff}.

The case for mixed dark matter from sterile neutrinos

Sterile neutrinos are $SU(2)$ singlets that mix with active neutrinos via a mass matrix,

its diagonalization leads to mass eigenstates that couple via standard model vertices. We study the cosmological production of heavy neutrinos via *standard model charged and neutral current vertices* under a minimal set of assumptions: i) the mass basis contains a *hierarchy of heavy neutrinos*, ii) these have very small mixing angles with the active (flavor) neutrinos, iii) standard model particles, including light (active-like) neutrinos are in thermal equilibrium. If kinematically allowed, the same weak interaction processes that produce active-like neutrinos also produce the heavier species. We introduce the quantum kinetic equations that describe their production, freeze out and decay and discuss the various processes that lead to their production in a wide range of temperatures assessing their feasibility as dark matter candidates. The final distribution function at freeze-out is a *mixture* of the result of the various production processes. We identify processes in which finite temperature *collective excitations* may lead to the production of the heavy species. As a specific example, we consider the production of heavy neutrinos in the mass range $M_h \lesssim 140\,\mathrm{MeV}$ from pion decay shortly after the QCD crossover including finite temperature corrections to the pion form factors and mass. We consider the different decay channels that allow for the production of heavy neutrinos showing that their frozen distribution functions exhibit effects from "kinematic entanglement" and argue for their viability as mixed dark matter candidates. We discuss abundance, phase space density and stability constraints and argue that heavy neutrinos with lifetime $\tau > 1/H_0$ freeze out of local thermal equilibrium, and *conjecture* that those with lifetimes $\tau \ll 1/H_0$ may undergo cascade decay into lighter DM candidates and/or inject non-LTE neutrinos into the cosmic neutrino background. A comparison is made between production through pion decays with the production of non-resonant production via active-sterile mixing.

Production of heavy sterile neutrinos from vector boson decay at $T \simeq M_W$.

We obtain a general quantum kinetic equation for production of sterile-like neutrinos and effective mixing angles in the medium valid in a wide range of temperature, to all orders in standard model interactions and to leading order for vacuum mixing angle $\theta \ll 1$. Effective mixing angles and production rates depend on helicity. We study production of sterile like

neutrinos in the standard model extended with a simple see-saw mass matrix in the temperature range $T \simeq M_w$ from vector boson decay in the medium. Helicity suppression entails that positive helicity massive sterile-neutrinos interact more weakly with the medium and their effective mixing angle is not substantially modified. Paradoxically, although negative helicity states are more profusely produced by vector boson decay, their mixing angle is strongly suppressed and their production rate effectively becomes smaller than that for the positive helicity component for sterile masses $M_s \lesssim 8.35\,\mathrm{MeV}$. We find an (MSW) resonance in absence of lepton asymmetry, which, however is screened by the damping rate and does not yield an enhancement in the production. Negative helicity sterile-like neutrinos freeze-out at $T_f^- \simeq 5\,\mathrm{GeV}$ whereas those for positive helicity neutrinos freeze-out at $T_f^+ \simeq 8\,\mathrm{GeV}$ with strongly non-thermal distributions. The distribution function of negative helicity states is broader than that for positive helicity stages as a result of the competition between decreasing rate and increasing mixing angle as temperature diminishes. While sterile neutrinos produced via vector boson decay do not simultaneously satisfy abundance, lifetime and cosmological constraints to be the sole dark matter component, massive sterile-like neutrinos are produced via this mechanism within the parameter range that recently was argued to solve the ^7Li problem, although with substantial tension in the parameter space. This production mechanism yields a heavier sterile like neutrino in the mass range few MeV, and could decay into lighter $\simeq$ KeV sterile neutrinos that could contribute to warm dark matter. We argue that heavy sterile neutrinos with lifetime $\leq 1/H_0$ reach local thermodynamic equilibrium.

BIBLIOGRAPHY

[1] E. Hubble, Proceedings of the National Academy of Sciences. 15 (3): 16873 (1929)

[2] ATLAS Collaboration, Physics Letters B, Vol 716, Issue 1, p.1 (2012)

[3] CMS Collaboration, Physics Letters B, Vol 716, Issue 1, p.30 (2012)

[4] S. L. Glashow, Nuclear Physics 22 (4): 579-588 (1961)

[5] S. Weinberg, Phys. Rev. Lett. 19 (21): 1264-1266 (1967)

[6] F. Englert, R. Brout, Phys Rev Lett 13 (9): 321-323 (1964)

[7] P. W. Higgs, Phys Rev Let, 13 (16): 508-509 (1964)

[8] Gunion, J.F.; Haber, H.E.; Kane, G.L.; Dawson, S. (1990). The Higgs hunters guide,. Addison-Wesley; J. Ellis, M. Gaillard, D. Nanopoulos, arXiv:1201.6045

[9] W. Hollik, Lectures given at the 2009 European School of High-Energy Physics; J. Gegelia, arXiv:1207.0156

[10] H. Fritzsch, M. Gell-Mann, Proceedings of the International Conference on Duality and Symmetry, Tel Aviv 1971, Proceedings of the Int. Conference on High Energy Physics, Chicago 1972; G. Zweig, CERN preprint 8182/TH401 (1964)

[11] L. Brown (1978), Physics Today 31 (9): 238 (1978)

[12] F. L. Wilson, American Journal of Physics. 36 (12): 1150. (1968)

[13] D. Griffiths, Introduction to Elementary Particles (2nd ed.), Wiley VCH (2009)

[14] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, R. P. Hudson, Physical Review. 105 (4): 14131415 (1957)

[15] R. L. Garwin, L. M. Lederman, M. Weinrich, Physical Review. 105 (4): 14151417 (1957)

[16] J. H. Christenson, J. W. Cronin, V. L. Fitch, R. Turlay, Physical Review Letters. 13:138 (1964)

[17] M. Peskin, D. Schroeder, An Introduction to Quantum Field Theory, Westview Press (1995)

[18] M. Srednicki, Quantum Field Theory, Cambridge University Press (2007)

[19] M. Markevitch; A. H. Gonzalez; D. Clowe; A. Vikhlinin; L. David; W. Forman; C. Jones; S. Murray, W. Tucker, Astrophys. J. 606 (2): 819824 (2003)

[20] S. Martin, arXiv:hep-ph/9709356; M. Dine, J. Mason, Reports on Progress in Physics, Volume 74, Number 5 (2011);

[21] E. Ponton, TASI lectures (2011), arXiv:1207.3827

[22] S. Liberati, L. Maccione, Ann. Rev. Nucl. Part. Sci. 59:245-267 (2009)

[23] K. Lane, TASI lectures (1993), arXiv:hep-ph/9401324

[24] S. Raby, arXiv:hep-ph/0608183

[25] T. Asaka, S. Blanchet, M. Shaposhnikov, Phys. Lett. B 631:151-156 (2005); T. Asaka, M. Shaposhnikov, Phys. Lett. B 620:17-26 (2005)

[26] C. W. Kim and A. Pevsner, *Neutrinos in Physics and Astrophysics*, (Harwood Academic Publishers, USA, 1993).

[27] S.M. Bilenky, C. Giunti, W. Grimus, Prog.Part.Nucl.Phys. **43**, 1 (1999)

[28] R. N. Mohapatra and P. B. Pal, *Massive Neutrinos in Physics and Astrophysics*, (World Scientific, Singapore, 2004).

[29] M. Fukugita and T. Yanagida, *Physics of Neutrinos and Applications to Astrophysics*, (Springer-Verlag Berlin Heidelberg 2003).

[30] L. Wolfenstein, Phys. Rev. **D 17**, 2369 (1978); S. P. Mikheyev and A. Yu. Smirnov, Sov. J. Nucl. Phys. **42**, 913 (1985).

[31] J. N. Bahcall, *Neutrino Astrophysics*, (Cambridge University Press, NY. 1989).

[32] W. C. Haxton, Ann. Rev. Astron. Astrophys. **33**, 459 (1995).

[33] W. Grimus, Lect. Notes Phys. **629**, 169 (2004).

[34] B. Kayser, arXiv:0804.1497; arXiv:0804.1121.

[35] R. N. Mohapatra *et al.*, Rept. Prog. Phys. **70**, 1757 (2007); R. N. Mohapatra and A. Y. Smirnov, Ann.Rev.Nucl.Part.Sci.**56**, 569 (2006).

[36] A. de Gouvea, Mod. Phys. Lett. A **19**, 2799 (2004).

[37] S. M. Bilenky, arXiv:1105.2306.

[38] C. Giunti, C. W. Kim, *Fundamentals of Neutrino Physics and Astrophysics*, (Oxford University Press, Oxford, 2007).

[39] W. C. Haxton,, arXiv:0808.0735.

[40] A. D. Dolgov, Phys. Rept. 370, 333 (2002), arXiv:hep-ph/0202122; Phys.Atom.Nucl.71, 2152 (2008), arXiv:0803.3887; Surveys High Energ.Phys.17, 91 (2002) arXiv:hep-ph/0208222.

[41] D. V. Forero, M. Trtola, J. W. F. Valle, Phys. Rev.**D86**, 073012 (2012).

[42] G. L. Fogli *et.al.* arXiv:1205.5254.

[43] A. Aguilar *et.al.* (LSND collaboration), Phys. Rev. **D64**, 112007 (2001); C. Athanassopoulos *et.al.* (LSND collaboration), Phys. Rev. Lett. **77**, 3082 (1996).

[44] A. A. Aguilar-Arevalo *et.al.* (MiniBooNE collaboration), Phys. Rev. Lett. **105**, 181801 (2010).

[45] A. A. Aguilar-Arevalo et al. [MiniBooNE Collaboration], arXiv:1207.4809; A. A. Aguilar-Arevalo et al. [MiniBooNE Collaboration], arXiv:1303.2588

[46] G. Cheng *et. al.* (MiniBooNE/SciBooNE Collaborations), arXiv:1208.0322.

[47] T. A. Mueller *et.al.*, Phys.Rev.**C83**, 054615 (2011).

[48] G. Mention *et.al* Phys. Rev. **D83**, 073006 (2011).

[49] M. Sorel, J. M. Conrad, M. Shaevitz, Phys. Rev. **D70**, 073004 (2004).

[50] G. Karagiorgi, A. Aguilar-Arevalo, J. Conrad, M. Shaevitz, K. Whisnant, M. Sorel, V. Barger, Phys. Rev. **D75**, 013011 (2007).

[51] G. Karagiorgi, Z. Djurcic, J. M. Conrad, M. H. Shaevitz, M. Sorel, Phys.Rev.**D80**, 073001 (2009).

[52] C. Giunti, M. Laveder, arXiv:1111.5211, Phys.Rev. **D84** (2011) 093006, Phys.Rev. **D84** (2011) 073008.

[53] G. Karagiorgi, Z. Djurcic, J. M. Conrad,M. H. Shaevitz, M. Sorel, Phys. Rev. **D80**,073001 (2009).

[54] M. Maltoni, T. Schwetz, Phys.Rev.**D76**, 093005 (2007).

[55] J. Kopp, M. Maltoni, T. Schwetz, Phys.Rev.Lett.**107**,091801 (2011); T. Schwetz, M. Tortola and J.W.F. Valle, New J. Phys. 10 (2008) 113011; C. Giunti, arXiv:1106.4479.

[56] C. Giunti, M. Laveder, Phys.Rev.**D82**, 093016 (2010); Phys.Rev.**D77**, 093002 (2008); Mod.Phys.Lett.**A22**, 2499 (2007).

[57] E. Akhmedov, T. Schwetz, JHEP **1010**, 115 (2010).

[58] C. Giunti, arXiv:1110.3914; C. Dib, J. C. Helo, S. Kovalenko, I. Schmidt, Phys.Rev. **D84**, 071301 (2011).

[59] J. M Conrad, C. M. Ignarra, G. Karagiorgi, M. H.Shaevitz, J. Spitz, arXiv:1207.4765.

[60] S. K. Agarwalla, *et. al.* (EUROnu-WP6 2010 Report), arXiv:1209.2825.

[61] G. Hinshaw (WMAP collaboration), arXiv:1212.5226.

[62] Z. Hou, *et al.*, (SPT collaboration) arXiv:1212.6267; J. L. Sievers, *et. al.*, (ACT collaboration) arXiv:1301.0824.

[63] P. Ade et al (Planck Collaboration), Astronomy and Astrophysics, Volume 571, A16, arXiv:1303.5076

[64] M. Archidiacono, N. Fornengo, C. Giunti, A. Melchiorri, arXiv:1207.6515

[65] A. Mirizzi, *et.al.* arXiv:1303.5368.

[66] T. L. Smith, S. Das, and O. Zahn, Phys.Rev. **D85** (2012); M. Archidiacono, E. Calabrese, and A. Melchiorri, Phys.Rev. **D84** (2011), 123008; K. M. Nollett and G. P. Holder, arXiv 1112.2683.

[67] J. Hamann, S. Hannestad, G. G. Raffelt, Y.Y.Y. Wong, JCAP **1109**, 034 (2011).

[68] K. Abazajian *et.al.* arXiv:arXiv:1204.5379; B. Kayser, arXiv:arXiv:1207.2167.

[69] S. Dodelson, L. M. Widrow, Phys. Rev. Lett. **72**, 17 (1994); S. Colombi, S. Dodelson, L. M. Widrow, Astrophys. J. **458**, 1 (1996).

[70] X. Shi, G. M. Fuller, Phys. Rev. Lett. **82**, 2832 (1999); K. Abazajian, G. M. Fuller, M. Patel, Phys. Rev. **D64**, 023501 (2001); K. Abazajian, G. M. Fuller, Phys. Rev. **D66**, 023526, (2002); G. M. Fuller et. al., Phys.Rev. **D68**, 103002 (2003); K. Abazajian, Phys. Rev. D73,063506 (2006).

[71] A. Kusenko, Phys.Rept.**481**, 1 (2009); Int.J.Mod.Phys.**D16**,2325 (2008); A. Kusenko, arXiv:hep-ph/0703116; arXiv:astro-ph/0608096; T. Asaka, M. Shaposhnikov, A. Kusenko; Phys.Lett. **B638**, 401 (2006); P. L. Biermann, A. Kusenko, Phys. Rev. Lett. **96**, 091301 (2006).

[72] A. Boyarsky, O. Ruchayskiy, M. Shaposhnikov, Ann.Rev.Nucl.Part.Sci.**59**, 191 (2009); M. Shaposhnikov, I. Tkachev, Phys. Lett. B639,414 (2006).

[73] H. J. de Vega, N. G. Sanchez, Mon.Not.Roy.Astron.Soc.**404** 885 (2010); H. J. de Vega, P. Salucci, N. G. Sanchez, New Astronomy, **17**, 653 (2012); H. J. de Vega, N. G. Sanchez arXiv:1109.3187; C. Destri, H. J. de Vega, N. G. Sanchez, arXiv:1204.3090.

[74] D. Boyanovsky, H. J. de Vega, N. Sanchez, Phys.Rev.**D77**,043518 (2008); D. Boyanovsky, J. Wu, Phys.Rev.**D83**,043524 (2011); D. Boyanovsky, Phys.Rev.**D78**, 103505 (2008); Phys.Rev.**D83**, 103504 (2011).

[75] I. M. Shoemaker, K. Petraki, A. Kusenko, JHEP **1009**, 060 (2010); S. Ando, A. Kusenko, Phys. Rev. **D 81**, 113006 (2010).

[76] A. Nucciotti (MARE Collaboration), Nucl.Phys.Proc.Suppl. **229** 155 (2012).

[77] H.J. de Vega, O. Moreno, E. Moya de Guerra, M. Ramon Medrano, N. Sanchez, Nucl. Phys. **B866** 177, (2013).

[78] M. Galeazzi *et.al.* arXiv:1202.4763 and references therein.

[79] See Project X website: http://projectx.fnal.gov/pdfs/ProjectXwhitepaperJan.v2.pdf.

[80] J. L. Hewett *et.al.* arXiv:1205.2671

[81] A. J. Anderson *et.al.* Phys.Rev. **D86**, 013004 (2012).

[82] J. Spitz, Phys.Rev. **D85**, 093020 (2012).

[83] P. Kyberd *et.al.* arXiv:1206.0294.

[84] A method to distinguish between Dirac and Majorana neutrinos has been discussed in R. Shrock and B. Kayser, Phys. Rev.Lett.**112B**, 137 (1982).

[85] D. Boyanovsky, Phys. Rev. **D 84**, 065001 (2011).

[86] D. Hernandez, A. Yu. Smirnov, Phys.Lett. **B706**, 360 (2012); E. Kh. Akhmedov, D. Hernandez, A. Yu. Smirnov, JHEP **1204**, 052 (2012).

[87] A. G. Cohen, S. L. Glashow, Z. Ligeti, Phys.Lett.**B678**, 191 (2009).

[88] A. de Gouvea, B. Kayser, R. Mohapatra, Phys.Rev. **D67**, 053004 (2003).

[89] R. E. Shrock, Phys. Rev. **D24**, 1232 (1981); Phys. Lett. **96B**, 159 (1980).

[90] R. Abela, *et.al.* Phys. Lett. **B105**, 263 (1981).

[91] F. P. Calaprice *et.al.* Phys. Lett. **B106**, 175 (1981).

[92] R. C. Minehart *et.al.* Phys. Rev. Lett. **52**, 804 (1984).

[93] M. Daum *et.al.* Phys. Rev. **D36**, 2624 (1987); M. Daum *et.al.*, Phys. Lett. **B361**, 179 (1995); Phys. Rev. Lett.**85**, 1815 (2000).

[94] R. Bilger *et.al.* Phys. Lett. **B363**, 41 (1995).

[95] P. Astier *et.al.* Phys. Lett. **B527**, 23 (2002).

[96] Y. Asano *et.al.* Phys. Lett. **B104**, 84 (1981).

[97] R. S. Hayano *et.al.* Phys. Rev. Lett. **49**, 1305 (1982).

[98] A. Kusenko, S. Pascoli, D. Semikoz, JHEP **0511**, 028 (2005).

[99] D. I. Britton, *et.al.* Phys. Rev. **D46**, R885 (1992); D. Britton, *et.al.*, Phys. Rev. Lett. **68**, 3000 (1992).

[100] V. Weisskopf, E. Wigner, Z. Phys. **63**, 54 (1930).

[101] P. K. Kabir, *The CP puzzle: strange decays of the neutral kaon*, (Academic Press, N.Y. 1968) (see appendix A).

[102] J. Wu, J. A. Hutasoit, D. Boyanovsky, R. Holman, Phys. Rev.**D82**, 013006 (2010); Int. J. Mod. Phys. **A 26**, 5261 (2011).

[103] T. Goldman, Mod.Phys.Lett. **A25** 479 (2010).

[104] E. Kh. Akhmedov, A. Yu. Smirnov, Found.Phys.**41**, 1279 (2011).

[105] M. Nauenberg, Phys. Lett. **B447**, 23 (1999).

[106] B. Meszena, A. Patkos, Mod.Phys.Lett.**A26**, 101 (2011).

[107] S. M. Bilenky, Phys.Part.Nucl.**41**, 690 (2010); S. M. Bilenky, C. Giunti, arXiv:1203.5250.

[108] D. Delepine, V. Gonzalez Macias, S. Khalil, G. Lopez Castro, Phys.Rev.**D79**,093003 (2009); AIP Conf.Proc.**1361**, 395 (2011); Phys.Lett. **B693**, 438 (2010); AIP Conf.Proc. **1361**, 279 (2011); J.Phys.Conf.Ser. **315**, 012026 (2011).

[109] See chapters 8.2 and 8.3 in ref.[38].

[110] B. Kayser, Phys. Rev. **D 24**, 110 (1981); Nucl. Phys. Proc. Suppl. **118**, 425 (2003).

[111] J. Rich, Phys. Rev. **D 48**, 4318 (1993).

[112] C. Giunti, JHEP **0211**, 017 (2002); J.Phys. **G34** R93 (2007); Found.Phys.Lett. **17** , 103 (2004); Mod.Phys.Lett. **A16** , 2363(2001); C. Giunti, C. W. Kim, Found.Phys.Lett. **14**, 213 (2001); Phys.Rev. **D58**, 017301 (1998); C. Giunti, C. W. Kim, U. W. Lee, Phys. Lett. **B421**, 237 (1998).

[113] A. D. Dolgov, O. V. Lychkovskiy, A. A. Mamonov, L. B. Okun and M. G. Schepkin, Eur. Phys. J. C **44**, 431 (2005); A. D. Dolgov, O. V. Lychkovskiy, A. A. Mamonov, L. B. Okun, M. V. Rotaev and M. G. Schepkin, Nucl. Phys. B **729**, 79 (2005); A. D. Dolgov, L. B. Okun, M. V. Rotaev and M. G. Schepkin, arXiv:hep-ph/0407189.

[114] M. Beuthe, Phys.Rept. **375**, 105 (2003).

[115] C. M. Ho, D. Boyanovsky, Phys.Rev. **D73**, 125014 (2006).

[116] E. K. Akhmedov and J. Kopp, JHEP **1004**, 008 (2010).

[117] For recent reviews and fits see: M. Gonzalez-Garcia, M. Maltoni, J. Salvado, JHEP **04**, 056 (2010); JHEP **1008**, 117 (2010); M.C. Gonzalez-Garcia, M. Maltoni, Phys.Rept.**460**,1 (2008); A. Strumia, F. Vissani, arXiv:hep-ph/0606054v3.

[118] G. J. Feldman, J. Hartnell, T. Kobayashi, arXiv:1210.1778

[119] B. Bhattacharya, A. M. Thalapillil, C. E. M. Wagner, Phys.Rev. **D85**, 073004 (2012).

[120] A. de Gouvea, T. Wytock, Phys. Rev. **D79**, 073005 (2009).

[121] L. Canetti, M. Drewes, T. Frossard, M. Shaposhnikov, Phys. Rev. **D 87**, 093006 (2013); M. Drewes, B. Garbrecht, JHEP **1303** 096 (2013); L. Canetti, M. Drewes, M. Shaposhnikov, New J. Phys. **14**, 095012 (2012); Phys. Rev. Lett. **110**, 061801 (2013).

[122] A. Bungau *et.al.* Phys. Rev. Lett. **109**, 141802 (2012).

[123] C. Kraus, A. Singer, K. Valerius, C. Winheimer, Eur. Phys. J. **C 73**, 2323 (2013).

[124] S. N. Gninenko, Phys. Rev. **D85**, 051702 (2012), Phys.Rev.**D83**, 093010 (2011); Phys.Rev.**D83**, 015015 (2011); Phys.Rev.Lett.**103**, 241802 (2009) ; S. N. Gninenko, D. S. Gorbunov, Phys.Rev.**D81**, 075013 (2010).

[125] M. Masip, P. Masjuan, D. Meloni, arXiv: 1210.1519.

[126] M. Aoki *et.al.* (PIENU collaboration), Phys.Rev. **D84**, 052002 (2011).

[127] S. Pakvasa, Lett. Nuovo Cim. **31**, 497 (1981).

[128] A. D. Dolgov, A. Y. Morozov, L. B. Okun and M. G. Shchepkin, Nucl. Phys. **B 502**, 3 (1997).

[129] C. Giunti and C. W. Kim, Found. Phys. Lett. **14**, 213 (2001).

[130] H. Burkhardt, J. Lowe, G. J. Stephenson and T. Goldman, Phys. Rev. **D 59**, 054018 (1999).

[131] E. Sassaroli, Y. N. Srivastava and A. Widom, Z. Phys. **C 66**, 601 (1995); Y. N. Srivastava and A. Widom, hep-ph/9707268; Y. Srivastava, A. Widom and E. Sassaroli, Eur. Phys. J. **C2**, 769 (1998).

[132] J. H. Field, Eur. Phys. J. **C 37**, 359 (2004); hep-ph/0110064; hep-ph/0110066; hep-ph/0303152.

[133] E. K. Akhmedov, JHEP**0709**, 116 (2007).

[134] T.E. Clark, S.T. Love, Mod.Phys.Lett. **A19**, 297 (2004).

[135] D. Boyanovsky, C. M. Ho, Astropart.Phys. **27**, 99 (2007).

[136] J. Adam *et.al.* (MEG collaboration), Nucl. Phys. **B834**, 1 (2010); Phys. Rev. Lett.**107**, 171801 (2011).

[137] H. Natori (for the MEG collaboration), Nucl.Phys.Proc.Suppl. **210-211**, 241 (2011); Y. Uchiyama (for the MEG collaboration), Proceedings of Science PoS **HQL2010**, 055 (2011).

[138] The Mu2e Project, Collaboration *Mu2e Conceptual Design Report*, arXiv:1211.7019.

[139] L. Lello, D. Boyanovsky, Phys. Rev. **D 87**, 073017 (2013).

[140] B. A. Kniehl, A. Sirlin, Phys.Rev. **D85**, 036007 (2012) 036007 ; B. A. Kniehl, arXiv:1308.3140.

[141] B. Machet, arXiv:1203.5902; Q. Duret, B. Machet, Phys. Lett. **B643**, 303 (2006); Q. Duret, B. Machet, M. I. Visotsky, Eur. Phys. J. **C61**, 247 (2009); Q. Duret, B. Machet, Mod. Phys. Lett. **A24**, 273 (2009).

[142] D. Boyanovsky, R. Holman, J. Hutasoit, Phys.Rev.**D81** 033009 (2010); D. Boyanovsky, R. Holman, J. Hutasoit, Phys.Rev.**D80**,025012 (2009).

[143] F. F. Deppisch, Fortsch.Phys. **61**, 622 (2013).

[144] A. Abada, Comptes Rendus Physique **13**,180 (2012), (arXiv:1110.6507).

[145] J. Chakraborty, P. Ghosh, W. Rodejohann, Phys. Rev. **D 86**, 075020 (2012).

[146] F. Renga, arXiv:1208.3593.

[147] J. C. Helo, S. Kovalenko, I. Schmidt, Nucl.Phys.**B853**, 80 (2011).

[148] A. Einstein, B. Podolsky, N. Rosen, Phys. Rev. **47**, 777 (1935).

[149] S. J. Van Enk, N. Lutkenhaus, H. J. Kimble, Phys. Rev. **A 75**, 052318 (2007).

[150] B. B. Blinov, D. L. Moehring, L. M. Duan, C. Monroe, Nature, **428**, 153 (2004); J. Volz *et.al.* Phys. Rev. Lett. **96**, 030404 (2006); T. Wilk, S. C. Webster, A. Kuhn, G. Rempe, Science **317**, 488 (2007).

[151] E. Togan *et.al.*, Nature *466*, 730, (2010); K. De Greve, *et.al.* Nature **491**, 421 (2012); W. B. Gao, *et.al.* Nature **491**,426 (2012); J. R. Schaibley *et.al.* arXiv:1210.5555.

[152] Nielsen M A and Chuang I L, *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press, 2000).

[153] J. Preskill, arXiv:1203.5813.

[154] R. Horodecki, P. Horodecki, M. Horodecki, K. Horodecki, Rev.Mod.Phys.**81**, 865 (2009).

[155] M. C. Tichy, F. Mintert, A. Buchleitner, J. Phys. B: At. Mol. Opt. Phys. **44** 192001 (2011).

[156] A. Go (for the Belle collaboration), Phys.Rev.Lett. **99**, 131802 (2007).

[157] J. P. Lees (BaBar Collaboration), Phys. Rev. Lett. **109**, 211801 (2012).

[158] Ray F. Cowan, *for the BABAR Collaboration*, arXiv:1301.1372

[159] J. Bernabeu, F. Martinez-Vidal, P. Villanueva-Perez, JHEP **1208**, 064 (2012).

[160] D. Atwood, A. Soni, Phys.Rev. **D82**, 036003 (2010).

[161] A. Kumar Alok, S. Banerjee, arXiv:1304.4063.

[162] D. Boyanovsky, Physical Review **D84**, 065001 (2011); L. Lello, D. Boyanovsky, Phys. Rev. **D 87**, 073017 (2013).

[163] P. Calabrese, J. Cardy, J.Stat.Mech.**0406**, P06002 (2004); Int.J.Quant.Inf. **4** , 429 (2006); J.Phys.**A42**, 504005 (2009).

[164] For a review see: J. Eisert, M. Cramer and M. B. Plenio, Rev. Mod. Phys. **82**, 277 (2010).

[165] C. G. Callan and F. Wilczek, Phys. Lett. B **333**, 55 (1994)

[166] M. Hertzberg, J. Phys. A: Math. Theor. **46**, 015402 (2013).

[167] M. Srednicki, Phys. Rev. Lett. **71**, 666 (1993).

[168] H. Casini, M. Huerta, J.Phys.**A42**, 504007 (2009); J.Stat.Mech.**0801**, P01012 (2008).

[169] S. N. Solodukhin, Living Rev. Rel. **14**, 8 (2011).

[170] T. Jacobson, A. Satz, PhysRev **D87**, 084047 (2013).

[171] S.-Y.Lin, C.-H. Chou, B. L. Hu, Phys. Rev. **D81**, 084018 (2010).

[172] J. Maldacena, G. L. Pimentel, JHEP **1302**, 038 (2013).

[173] V. Balasubramanian, M. B. McDermott, M. Van Raamsdonk, Phys. Rev. **D86**, 045014 (2012); T.-C. L. Hsu, M. B. McDermott, M.Van Raamsdonk, arXiv:1210.0054.

[174] D. Boyanovsky, Phys. Rev. **A 87**, 033815 (2013).

[175] D. Boyanovsky, R. Holman, JHEP, Volume 2011, Number 5, 47 (2011); D. Boyanovsky, Phys. Rev. D 85, 123525 (2012); Phys. Rev. D 86, 023509 (2012).

[176] D. Boyanovsky, H. J. de Vega, Annals Phys. **307**, 335 (2003); D. Boyanovsky, H. J. de Vega, S.-Y. Wang, Phys.Rev. **D67**, 065022 (2003.

[177] A. M. Polyakov, Sov. Phys. Usp. **25** , 187 (1982) [Usp. Fiz. Nauk **136**, 538 (1982)].

[178] N. P. Myhrvold, Phys. Lett. **B 132**, (308) (1983); Phys. Rev. **D 28**, 2439 (1983).

[179] L. H. Ford, Phys. Rev. **D 31**, 710 (1985) 710.

[180] B. Allen, Phys. Rev. **D 32**, 3136, (1985); B. Allen, A. Folacci, Phys. Rev. **D35**, 3771 (1987).

[181] A. Folacci, Phys. Rev. **D 46**, 2553, (1992).

[182] , A. D. Dolgov, M. B. Einhorn, V. I. Zakharov, Phys. Rev. **D52**, 717 (1995); Acta Phys.Polon. **B26** , 65 (1995); A. Dolgov, D. N. Pelliccia, Nucl. Phys. **B734**, 208 (2006).

[183] S. P. Miao, N. C. Tsamis, R. P. Woodard, J.Math.Phys.**51**, 072503 (2010); R. P. Woodard, arXiv:astro-ph/0310757; T. M. Janssen, S. P. Miao, T. Prokopec, R. P. Woodard, Class.Quant.Grav.**25**, 245013 (2008); N. C. Tsamis and R. P. Woodard, Phys. Lett. **B 301**, 351 (1993) 351; N. C. Tsamis and R. P. Woodard, Annals Phys. **238**,1 (1995); N. C. Tsamis and R. P. Woodard, Phys. Rev. **D 78**, 028501 (2008).

[184] I. Antoniadis, P. O. Mazur, E. Mottola, NewJ.Phys.**9**, 11 (2007); E. Mottola, R. Vaulin, Phys. Rev. **D74** , 064004 (2006); M. Giannotti, E. Mottola, Phys. Rev.**D79**,045014 (2009); E. Mottola, Acta Physica Polonica **B**, 2031 (2010) Vol.41, issue 9; Int.J.Mod.Phys.**A25**, 2391 (2010).

[185] I. Antoniadis, J. Iliopoulos and T. N. Tomaras, Phys. Rev. Lett. **56**, 1319 (1986); I. Antoniadis and E. Mottola, J. Math. Phys. **32**, 1037 (1991).

[186] L. R. W. Abramo, R. H. Brandenberger and V. F. Mukhanov, Phys. Rev. **D 56**, 3248 (1997); V. F. Mukhanov, L. R. W. Abramo and R. H. Brandenberger, Phys. Rev. Lett. **78**, 1624 (1997); L. R. W. Abramo and R. P. Woodard, Phys. Rev. **D 60**, 044010 (1999); L. R. Abramo and R. P. Woodard, Phys. Rev. **D 65**, 063515 (2002); G. Geshnizjani and R. Brandenberger, JCAP **0504**, 006 (2005); R. H. Brandenberger, arXiv:hep-th/0210165.

[187] A. M. Polyakov, Nucl.Phys.**B797**, 199 (2008); Nucl.Phys.**B834**,316 (2010); arXiv:1209.4135; D. Krotov, A. M. Polyakov, Nucl.Phys.**B 849** 410 (2011).

[188] P. R. Anderson, E. Mottola, arXiv:1310.0030; arXiv:1310.1963.

[189] D. Boyanovsky, H. J. de Vega, Phys. Rev. **D70**, 063508 (2004); D. Boyanovsky, H. J. de Vega, N. G. Sanchez, Phys. Rev.**D71** 023509 (2005); Nucl. Phys. **B747**, 25 (2006).

[190] D. Boyanovsky, R. Holman, S. Prem Kumar, Phys. Rev. **D56**, 1958 (1997).

[191] J. Bros, H. Epstein, M. Gaudin, U. Moschella and V. Pasquier, Commun. Math. Phys. **295**, 261 (2010); J. Bros, H. Epstein and U. Moschella, Annales Henri Poincare **11** 611 (2010; J. Bros, H. Epstein and U. Moschella, JCAP **0802**, 003 (2008).

[192] E. T. Akhmedov, A. Roura, A. Sadofyev, Phys. Rev.**D82**, 044035 (2010); E. T. Akhmedov, P. V. Buividovich, Phys. Rev.**D78**, 104005 (2008); E. T. Akhmedov, Mod.Phys.Lett.**A25**,2815 (2010); E. T. Akhmedov, P. V. Buividovich, D. A. Singleton, Phys.Atom.Nucl. **75**, 525 (2012) 525; E. T. Akhmedov, JHEP **1201** 066 (2012); E. T. Akhmedov, Ph. Burda Phys.Rev. **D86** (2012) 044031; E. T. Akhmedov, Phys.Rev. D87 (2013) 044049.

[193] E.T. Akhmedov, Int. J. of Mod. Phys. **D23**, No. 1, 1430001 (2014).

[194] D. Marolf, I. A. Morrison, M.Srednicki, Class. Quant. Grav. **30**, 155023 (2013).

[195] M. Giovannini, Phys.Rev.**D83**,023515 (2011).

[196] L. Amico, R. Fazio, A. Osterloh and V. Vedral, Rev. Mod. Phys. **80**, 517 (2008).

[197] S. Das, S. Shankaranarayanan, S. Sur, arXiv:0806.0402.

[198] L.Lello, D. Boyanovsky, R. Holman, JHEP **2013**,116 (2013).

[199] C.P. Burgess, R. Holman, L. Leblond, S. Shandera JCAP **1003**, 033 (2010); JCAP **1010**, 017 (2010).

[200] B. Garbrecht, G. Rigopoulos, Phys. Rev. **D 84**, 063516 (2011).

[201] D. P. Jatkar, L. Leblond, A. Rajaraman, Phys.Rev. **D85** (2012) 024047.

[202] M. van der Meulen, J. Smit, JCAP **0711**, 023 (2007).

[203] D. Mazur, J. S. Heyl, Phys.Rev.**D80**, (2009) 023523.

[204] M. O. Scully, M. S. Zubairy, *Quantum Optics*, (Cambridge University Press, Cambridge, U.K. (1997)); M. Sargent III, M. Scully, W. E. Lamb, Laser Physics (Addison-Wesley, Reading MA 1974); W. Louisell, Quantum Statistical Properties of Radiation, (Wiley, N.Y. 1974).

[205] H. V. Peiris *et.al.* Astrophys. J. Suppl. **148**, 213 (2003).

[206] N. Bartolo, E. Komatsu, S. Matarrese, A. Riotto, Phys.Rept. **402**, 103 (2004).

[207] J. M. Maldacena, JHEP **0305**, 013, (2003).

[208] C. Armendariz-Picon, JCAP **0702**, 031 (2007).

[209] D. Larson, *et.al.* Astrophys.J.Suppl.192:16,2011.

[210] G. Hinshaw, *et.al.* (WMAP collaboration), Astrophys.J.Suppl. **208**, 19.

[211] A. Gruppuso, P. Natoli, F. Paci, F. Finelli, D. Molinari, A. De Rosa, N. Mandolesi, JCAP **1307**, 047 (2013).

[212] B. R. Holstein, *Topics in Advanced Quantum Mechanics*, (Addison-Wesley, Redwood City, CA, 1992).

[213] A. B. Carter, A. I. Sanda, Phys. Rev. **D23**, 1567 (1981).

[214] Ya. I. Azimov, JETP Lett. **50**, 447 (1989); Sov. Phys.JETP **71**, 31 (1990); Phys. Rev. **D42**, 3705 (1990).

[215] Y. Azimov, I. Dunietz, Phys.Lett.**B395**, 334 (1997).

[216] B. Kayser, L. Stodolsky, Phys. Lett. **B359**, 343 (1995); B. Kayser, L. Stodolsky, arXiv:hep-ph/9610522, B. Kayser, arXiv:hep-ph/9709382.

[217] A. Amorim, M. G. Santos, J. P. Silva, Phys. Rev. **D59**, 056001 (1999).

[218] W. Grimus, L. Lavoura, Phys.Lett. **B387**, 195 (1996); G. C. Branco, W. Grimus, L. Lavoura, Phys.Lett. **B372**, 311 (1996).

[219] L. Lavoura, Phys.Rev. **D62** (2000) 056002; L. Lavoura, J. P. Silva, Phys.Rev. **D60** (1999) 056003; G. C. Branco, L. Lavoura, J. P. Silva *CP violation* (Oxford University Press, N.Y. 1999).

[220] M. A. Gomshi Nobary, B. Mojaveri, Phys. Lett.**B649**, 417 (2007).

[221] B.D. Yabsley, arXiv:0810.1822.

[222] G. Cvetic, C. S. Kim, J. Zamora-Saa, arXiv:1311.7554; G. Cvetic, C. Dib, C. S. Kim, JHEP **1206**,149 (2012); C. Dib, C. S. Kim, arXiv:1403.1985; G. Cvetic, C. Dib, S. K. Kang, C. S. Kim, Phys. Rev. **D82**, 053010 (2010).

[223] Y. Grossman, M. Martone, D. J. Robinson, JHEP **10**, 127 (2011), and references therein.

[224] G. S. Agarwal, *Quantum Optics*, (Cambridge University Press, N.Y. 2013).

[225] B.J.Dalton, B.M.Garraway, Phys. Rev. **A 68**, 033809 (2003); B.M. Garraway, B.J. Dalton, J. Phys. **B 39**, S767 (2006).

[226] F. Troiani, J. I. Perea, and C. Tejedor, Phys. Rev. **B 74**, 235310 (2006); F. Troiani, C. Tejedor, Phys. Rev. **B 78**, 155305 (2008).

[227] A. Greilich, M. Schwab, T. Berstermann, T. Auer, R. Oulton, D. R. Yakovlev, M. Bayer, V. Stavarache, D. Reuter, and A. Wieck, Phys. Rev. **B 73**, 045323, (2006).

[228] T. Prokopec and R. P. Woodard, JHEP **0310**, 059 (2003).

[229] D. Boyanovsky, H. J. de Vega, Phys. Rev. **D70**, 063508 (2004); D. Boyanovsky, H. J. de Vega, N. G. Sanchez, Phys. Rev.**D71** 023509 (2005); Nucl. Phys. **B747**, 25 (2006); D. Boyanovsky, R. Holman, S. Prem Kumar, Phys. Rev. **D56**, 1958 (1997).

[230] Q. Wang, A. I. Sanda, Phys. Rev. **D55**, 3131, (1997).

[231] L. Lello, D. Boyanovsky, Phys. Rev. **D 87**, 073017 (2013); L. Lello, D. Boyanovsky, R. Holman, JHEP **2013**,116 (2013).

[232] A. Muthukrishnan, M. O. Scully, M. Suhail Zubairy, J. Opt. B: Quantum Smiclass. Opt. **6**, S575 (2004).

[233] B. Garbrecht and M. Herranen, Nucl. Phys. **B 861**, 17 (2012).

[234] M. Garny, A. Kartavtsev and A. Hohenegger, Annals Phys. **328**, 26 (2013).

[235] T. Frossard, M. Garny, A. Hohenegger, A. Kartavtsev and D. Mitrouskas, "Systematic approach to thermal leptogenesis," Phys. Rev. **D 87** 085009 (2013).

[236] A. Anisimov, W. Buchmller, M. Drewes and S. Mendizabal, Annals Phys. **326**, 1988 (2011), Erratum-ibid. **338**, 376 (2011) 376.

[237] A. A. Starobinsky, JETP Lett. 30, 682 (1980); Phys. Lett. **91B**, 99 (1980); V. F. Mukhanov, G. V. Chibisov, Soviet Phys. JETP Lett. 33, 532 (1981).

[238] A. H. Guth, Phys. Rev. **D23**, 347 (1981).

[239] A. A. Linde, Phys. Lett. **108B**, 389 (1982); Phys. Lett. **116B**,335 (1982); Phys. Lett. **129B**,177 (1983).

[240] A. A. Albrecht and P. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).

[241] V. F. Mukhanov, H. A. Feldman , R. H. Brandenberger, Phys. Rept. 215, 203 (1992).

[242] A. Riotto, arXiv: hep-ph/0210162.

[243] D. Baumann, arXiv:0907.5424.

[244] M. Giovannini, Int. J. Mod. Phys. D14 363 (2005).

[245] J. Lidsey, A. R. Liddle, E. Kolb, Rev. of Mod. Phys. **69**, 373 (1997).

[246] E. Komatsu *et.al.* (WMAP collaboration), Astrophys.J.Suppl.**192**, 18 (2011).

[247] G. Hinshaw, A. J. Branday, C. L. Bennett, *et al.* , ApJ, **464**, L25 (1996).

[248] Bond, J. R., Jaffe, A. H., Knox, L. Phys. Rev. **D 57**, 2117 (1998).

[249] D. N. Spergel, **et al.** (WMAP collaboration), ApJS, **148**, 175 (2003).

[250] C. L. Francis and J. A. Peacock, MNRAS **406**, 14 (2010).

[251] A. Rassat, J.-L. Starck, F.-X. Dupe, arXiv:1303.4727.

[252] C. J. Copi, D. Huterer, D. J. Schwarz, G. D. Starkman, Advances in Astronomy vol. 2010, Article ID 847541 (2010); C. J. Copi, D. Huterer, D. J. Schwarz, G. D. Starkman, arXiv:1310.3831; A. Yoho, C. J. Copi, G. D. Starkman, A. Kosowsky, arXiv:1310.7603.

[253] T. S. Bunch and P. C. Davies, Proc. R. Soc. A360, 117 (1978); N. D. Birrell and P. C. W. Davies, Quantum elds in curved space, (Cambridge Monographs in Mathematical Physics, Cambridge University Press, Cambridge, 1982).

[254] N. Kaloper, M. Kleban, A. Lawrence, S. Shenker and L. Susskind, JHEP **0211**, 037 (2002).

[255] B. Greene, K. Schalm, J. P. van der Schaar and G. Shiu, In the Proceedings of 22nd Texas Symposium on Relativistic Astrophysics at Stanford University, Stanford, California, 13-17 Dec 2004, pp 0001 [arXiv:astro-ph/0503458].

[256] R. Easther, W. H. Kinney and H. Peiris, JCAP **0508**, 001 (2005).

[257] R. Brunetti, K. Fredenhagen and S. Hollands, JHEP **0505**, 063 (2005).

[258] K. Goldstein and D. A. Lowe, Nucl. Phys. **B 669**, 325 (2003).

[259] H. Collins and R. Holman, Phys. Rev. **D 70**, 084019 (2004).

[260] H. Collins, R. Holman and M. R. Martin, Phys. Rev. **D 68**, 124012 (2003); C. P. Burgess, J. M. Cline, F. Lemieux and R. Holman, JHEP 0302, 048 (2003); C. P. Burgess, J. M. Cline and R. Holman, JCAP 0310, 004 (2003).

[261] J. Martin and R. Brandenberger, Phys. Rev. D 68, 063513; R. H. Brandenberger and J. Martin, Int. J. Mod. Phys. A 17, 3663 (2002).

[262] U. H. Danielsson, Phys. Rev. D 66, 023511 (2002); U. H. Danielsson, JHEP 0207, 040 (2002).

[263] A. Berera, L.-Z. Fang, G. Hinshaw, Phys.Rev. **D57**, 2207 (1998).

[264] A. Berera, A. F. Heavens, Phys.Rev.**D62** 123513 (2000).

[265] R. Holman, Andrew J. Tolley, JCAP **0805**, 001 (2008).

[266] Nishant Agarwal, R. Holman, Andrew J. Tolley, Jennifer Lin, JHEP **1305**, 085 (2013).

[267] J. Ganc, Phys. Rev. **D 84**, 063514 (2011).

[268] I. Agullo, J. Navarro-Salas, L. Parker, JCAP **1205**, 019 (2012).

[269] I. Agullo, L. Parker, Phys.Rev.**D83**, 063526 (2011).

[270] R. Flauger, D. Green, R. A. Porto, arXiv:1303.1430.

[271] A. Aravind, D. Lorshbough, S. Paban, JHEP **1307**, 076 (2013).

[272] J. Ganc, E. Komatsu, Phys. Rev. **D86**, 023518 (2012).

[273] A. Linde, JHEP **11**, 052 (2001).

[274] C. Contaldi, M. Peloso, L. Kofman, A. Linde, JCAP **0307**, 002 (2003).

[275] D. Boyanovsky, H. J. de Vega, N. G. Sanchez, Phys. Rev. **D74**, 123006, 123007 (2006).

[276] C. Destri, H. J. de Vega, N. G. Sanchez, Phys.Rev.**D81**063520 (2010); C. Destri, H. J. de Vega, N. G. Sanchez, Phys.Rev.**D78**, 023013 (2008); F. J. Cao, H. J. de Vega, N. G. Sanchez, Phys.Rev.**D78**, 083508 (2008).

[277] D. Boyanovsky, C. Destri, H. J. de Vega, N. G. Sanchez, Int.J.Mod.Phys.**A24**, 3669 (2009).

[278] M. M. Glenz, L. Parker, Phys.Rev.**D80**, 063534 (2009); M. M. Glenz, arXiv:0905.2641.

[279] E. Ramirez, D. J. Schwarz, Phys.Rev.**D80**, 023525 (2009); Phys. Rev. **D 85**, 103516 (2012).

[280] A. Ashoorioon, K. Dimopoulos, M. M. Sheikh-Jabbari, G. Shiu, arXiv:1306.4914; A. Ashoorioon, G. Shiu, JCAP **1103**, 025 (2011); A. Ashoorioon, A. Krause, arXiV: hep-th/0607001; A. Ashoorioon, A. Krause, K. Turzynski, JCAP **0902**, 014 (2009).

[281] S. Kundu, JCAP **1202**, 005 (2012).

[282] R. K. Jain, P. Chingangbam, J.-O. Gong, L. Sriramkumar and T. Souradeep, JCAP **0901**, 009 (2009); R. K. Jain, P. Chingangbam, L. Sriramkumar and T. Souradeep, Phys. Rev. **D 82**, 023509 (2010).

[283] L. Lello, D. Boyanovsky, JCAP 05, 029 (2014)

[284] S. Weinberg, Phys. Rev.**D72**, 043514 (2005); Phys. Rev. **D74**, 023508 (2006).

[285] D. Seery, Class. Quant. Grav. **27**, 124005 (2010); JCAP **0905**, 021 (2009); JCAP **0802**, 006 (2008); JCAP **0711**, 025 (2007).

[286] W. Xue, X. Gao, R. Brandenberger, JCAP **1206**, 035 (2012).

[287] S. B. Giddings, M. S. Sloth, JCAP **1101**, 023 (2011).

[288] C. T. Byrnes, M. Gerstenlauer, A. Hebecker, S. Nurmi, G. Tasinato, JCAP **1008**, 006 (2010); M. Gerstenlauer, A. Hebecker, G. Tasinato, JCAP **1106**, 021 (2011).

[289] W. Xue, K. Dasgupta, R. Brandenberger, Phys.Rev.**D83**, 083520 (2011).

[290] A. Rajaraman, Phys.Rev.**D82**, 123522 (2010).

[291] A. Riotto and M. S. Sloth, JCAP **0804**, 030 (2008).

[292] K. Enqvist, S. Nurmi, D. Podolsky, G. I. Rigopoulos, JCAP **0804**, 025 (2008).

[293] J. Serreau, R. Parentani, Phys.Rev. **D87**, 085012 (2013); R. Parentani, J. Serreau, Phys.Rev. **D87**, 045020 (2013) ; F. Gautier, J. Serreau, arXiv:1305.5705.

[294] A. A. Starobinski, J. Yokoyama, Phys. Rev. D50, 6357 (1994).

[295] R. P. Woodard, J.Phys.Conf.Ser.**68**, 012032 (2007); S.-P. Miao, R. P. Woodard ; Phys.Rev.**D74**, 044019 (2006); R. P. Woodard, arXiv:astro-ph/0502556; T. Brunier, V. K. Onemli, R. P. Woodard, Class.Quant.Grav.**22**, 59 (2005); E. O. Kahya and V. K. Onemli, Phys. Rev. **D76**, 043512 (2007); T. Prokopec, O. Tornkvist, R. Woodard, Phys.Rev.Lett.**89**, 101301 (2002).

[296] B. Garbrecht, T. Prokopec; Phys.Rev. **D73** 064036 (2006).

[297] T. Arai, Class. Quantum Grav. **29**, 215014 (2012).

[298] J. Serreau, Phys.Rev.Lett. **107**, 191103 (2011).

[299] L. Lello, D. Boyanovsky, R. Holman, arXiv:1305.2441.

[300] P. A. R. Ade *et.al.* (PLANCK collaboration), arXiv:1303.5075.

[301] P. A. R. Ade *et.al.* (PLANCK collaboration), arXiv:1303.5083.

[302] P. A. R. Ade *et.al.* (BICEP2 collaboration), arXiv:1403.3985.

[303] S. Das, A.Souradeep, JCAP **1402**, 002 (2014).

[304] D. Polarski, A. A. Starobinsky, Nucl. Phys. **B 385**, 623 (1992)

[305] P. D. Meerburg, J. P. van der Schaar, P. S. Corasaniti, JCAP **0905**, 018 (2009); P. . Meerburg, J. P. van der Schaar, M. G. Jackson, JCAP **1002**, 001 (2010); P. D. Meerburg Phys.Rev. **D82**, 063517 (2010).

[306] G. Nicholson, C. R. Contaldi, JCAP **01**, 002 (2008).

[307] W. H. Kinney, Phys.Rev. **D56** 2002 (1997).

[308] W. H. Kinney, Phys.Rev. **D72** 023515 (2005).

[309] N. C. Tsamis, R. P. Woodard Class. Quant. Grav. 21:93-102 (2003).

[310] N. C. Tsamis, R. P. Woodard, Phys. Rev. **D69**, 084005 (2004).

[311] L. Lello, D. Boyanovsky, R. Holman, Phys. Rev. **D 89**, 063533 (2014) .

[312] W.J. Handley, S.D. Brechet, A.N. Lasenby, and M.P. Hobson, Phys. Rev. **D89**, 063505 (2014).

[313] S. Choudhury, A. Mazumdar, Nucl. Phys. **B 882**, 386 (2014); S. Choudhury, A. Mazumdar, S. Pal, JCAP **07**, 041 (2013); S. Hotchkiss, A. Mazumdar, S. Nadathur, JCAP **02**, 008 (2012).

[314] C. R. Contaldi, M. Peloso, L. Sorbo, arXiv:1403.4596.

[315] P. D. Meerburg, D. N. Spergel, arXiv:1308.3705; P. D. Meerburg, D. N. Spergel, B. D. Wandelt, arXiv:1308.3704; P. D. Meerburg, R. Wijers, J. P. van der Schaar, arXiv:1109.5264.

[316] D. Polarski, A. A. Starobinsky, Phys.Lett. B356 (1995) 196-204.

[317] S. Dodelson, *Modern Cosmology* (Academic Press, N.Y. 2003); J. Lesgourgues, G. Mangano, G. Miel, S. Pastor, *Neutrino Cosmology*, Cambridge University Press (2013)

[318] P. Cushman et al. (Snowmass Summary) arXiv:1310.8327

[319] J. Kim, arXiv:0909.3908

[320] A. Boyarsky, D. Iakubovskyi, O. Ruchayskiy, Physics of the Dark Universe, Vol 1, issue 1, p.136-154 (2012), arXiv:1306.4954; A. Boyarsky, O. Ruchayskiy, M. Shaposhnikov, Ann. Rev. Nucl. Part. Sci. 59:191-214 (2009), arXiv:0901.0011

[321] W. J. G. de Blok, arXiv:0910.3538

[322] J. Dalcanton, C. Hogan, Astrophys.J. 561 (2001) 35-45, arXiv:astro-ph/0004381; C. Hogan, J. Dalcanton, Phys.Rev. D62 (2000) 063511, arXiv:astro-ph/0002330

[323] M. Boylan-Kolchin, J. Bullock, M. Kaplinghat, Mon. Not. Roy. Astron. Soc. 412:L40 (2011), arXiv:1103.0007; S Garrison-Kimmel, M. Boylan-Kolchin, J. Bullock, E. Kirby, arXiv:1404.5313; E. Papastergis, R. Giovanelli, M. Haynes, F. Shankar, arXiv:1407.4665

[324] M. G. Walker, J. Penarrubia, ApJ 742 20 2011, arXiv:1108.2404

[325] A. V. Tikhonov, A. Klypin, MNRAS 395, 1915 (2009), arXiv:0807.0924

[326] A. V. Tikhonov, S. Gottlober, G. Yepes, Y. Hoffman, MNRAS, vol 399, issue 3, pp. 1611-1621 (2009), arXiv:0904.0175

[327] B. Moore, T. Quinn, R. Governato, J. Stadel, G. Lake, Mon. Not. Roy. Astron. Soc. 310:1147-1152, 1999, arXiv:astro-ph/9903164

[328] P. Bode, J. P. Ostriker, N. Turok, Astrophys. J. 556 (2001) 93-107, arXiv:astro-ph/0010389

[329] V. Avila-Reese, P. Colin, O. Valenzuela, E. D'Onghia, C. Firmani, Astrophys. J. 559:516-530, 2001, arXiv:astro-ph/0010525

[330] M Lovell, C Frenk, V. Eke, A. Jenkins, L Gao, T Theuns, MNRAS 439 (1): 300-317 (2014), arXiv:1308.1399

[331] S. Colombi, S. Dodelson, L. M. Widrow, Astrophys. J. 458:1, 1996, arXiv:astro-ph/9505029

[332] K. Abazajian et al, arXiv:1204.5379

[333] M. Kaplinghat Phys. Rev. D72, 063510 (2005), arXiv:astro-ph/0507300; J. Cline, Y. Farzan. Z. Liu. G. Moore, W. Xue, Phys. Rev. D 89, 121302 (2014), arXiv:1404.3729

[334] K.A. Olive et al. (Particle Data Group), Chin. Phys. C, 38, 090001 (2014)

[335] A. Balantekin, W. Haxton, arXiv:1303.2272

[336] H. Minakata, arXiv:1403.3276

[337] F. P. An et al. (Daya Bay Collaboration) PRL 113, 141802 (2014); I. Girardi, D. Meloni, T. Ohlsson, H. Zhang, S. Zhou, JHEP 08 (2014) 057, arXiv:1405.6540

[338] D. Boyanovsky, L. Lello, New Journal of Physics 16 (2014) 063050, arXiv:1403.6366 ; D. Boyanovsky, Nucl. Phys. B888, 248 (2014), arXiv:1406.5739; D. Boyanovsky, arXiv:1409.4265

[339] L. Lello, D. Boyanovsky Nuclear Physics B, 880 (2014), arXiv:1212.4167

[340] T. Lasserre, conference proceedings, arXiv:1404.7352

[341] K. Abe, et al. (Super-Kamiokande Collaboration), arXiv:1410.2008

[342] A. Mirizzi, N. Saviano, G. Miele, P. D. Serpico, Phys.Rev. D86 053009, (2012) arXiv:1206.1046

[343] C. Giunti, arXiv:1311.1335

[344] M. Shaposhnikov, I. Tkachev Phys. Lett. B 639, 414-417 (2006), arXiv:hep-ph/0604236

[345] S. Gariazzo, C. Giunti, M. Laveder, arXiv:1404.6160

[346] R. Barbieri, A.D. Dolgov, Nucl. Phys. B349 (1991) 743-753; R. Barbieri, A.D. Dolgov, Phys. Lett. B237 (1990) 440-445; K. Kainulainen, Phys. Lett. B244 (1990) 191-195; K. Enqvist, K. Kainulainen, J. Maalampi, Phys. Lett. B249 (1990) 531-534; K. Enqvist, K. Kainulainen, J. Maalampi, Nucl. Phys. B349 (1991) 754-790; K. Enqvist, K. Kainulainen, M. Thomson, Nucl. Phys. B373 (1992)

[347] X. Shi, G. Fuller, Phys. Rev. Lett. 82, 2832-2835 (1999), arXiv:astro-ph/9810076

[348] D. Boyanovsky, Phys. Rev. D78, 103505 (2008), arXiv:0807.0646

[349] K. Petraki, A. Kusenko, Phys. Rev. D77, 065014 (2008), arXiv:0711.4646

[350] K. Petraki, Phys. Rev. D77, 105004 (2008), arXiv:0801.3470

[351] E. Bulbul, M, Markevitch, A. Foster, R. Smith, M. Loewenstein, S. Randall, ApJ 789 13 (2014), arXiv:1402.2301

[352] A. Boyarsky, O. Ruchayskiy, D. Iakubovskyi, J. Franse, arXiv:1402.4119; A. Boyarsky, J. Franse, D. Iakubovskyi, O. Ruchayskiy, arXiv:1408.2503

[353] T. Jeltema, S. Profumo, MNRAS 450, 2143 (2015) arXiv:1408.1699; T. Jeltema, S. Profumo, arXiv:1411.1759

[354] D. Malyshev, A. Neronov, D. Eckert, Phys. Rev. D 90, 103506 (2014) arXiv:1408.3531

[355] K. Abazajian, Phys. Rev. Lett. 112, 161303 (2014), arXiv:1403.0954

[356] A. Merle, Int.J.Mod.Phys. D22 1330020 (2013), arXiv:1302.2625; A. Merle, A. Schneider, arXiv:1409.6311; A. Merle, M. Totzauer, JCAP06, 011 (2015), arXiv:1502.01011; A. Merle, V. Niro, D. Schmidt, JCAP 1403, 028 (2014), arXiv:1306.3996.

[357] S. Horiuchi, P. Humphrey, J. Onorbe, K. Abazajian, M. Kaplinghat, S. Garrison-Kimmel, Phys. Rev. D89, 025017 (2014), arXiv:1311.0282

[358] M. Kaplinghat, R. Lopez, S. Dodelson, R. Scherrer, Phys. Rev. D60, 123508 (1999), arXiv:astro-ph/9907388

[359] C. Destri, H. J. de Vega, N. Sanchez, Astroparticle Physics, 46, 14 (2013), arXiv:1301.1864; C. Destri, H. J. de Vega, N. Sanchez, Phys. Rev. D 88, 083512 (2013), arXiv:1308.1109; D. Boyanovsky, H. J. de Vega, N. Sanchez, Phys. Rev. D78, 063546 (2008), arXiv:0807.0622; H. J. de Vega, P. Salucci, N. Sanchez, New Astron. 17 (2012) 653-666, arXiv:1004.1908.

[360] D. Kirilova, arXiv:1407.1784

[361] G. Steigman, Phys. Rev. D 87 (2013) 103517, arXiv:1303.0049; G. Steigman, Advances in High Energy Physics, 2012 (2012) 268321 arXiv:1208.0032

[362] M. Archidiacono, E. Giusarma, S. Hannestad, O. Mena, Advances in High Energy Physics, arXiv:1307.0637

[363] G. M. Fuller, C. T. Kishimoto, A. Kusenko, arXiv:1110.6479; A. V. Patwardhan, G. M. Fuller, C. T. Kishimoto, A. Kusenko, Phys. Rev. D 92, 103509 (2015), arXiv:1507.01977.

[364] C. Boehm, M. Dolan, C. McCabe, JCAP 1212 (2012) 027, arXiv:1207.0497

[365] J. Hasenkamp, arXiv:1405.6736

[366] J. Birrell, C. Yang, P. Chen, J. Rafelski, Phys. Rev. D 89, 023008 (2014), arXiv:1212.6943; J. Birrell, J. Rafeslki, Phys. Lett. B, 741, 77 (2015) arXiv:1404.6005

[367] J. Birrell, J. Rafelski, arXiv:1404.6005

[368] K. Fukushima, T. Hatsuda, 2011 Rep. Prog. Phys. 74 014001, arXiv:1005.4814

[369] T. Bhattacharya et al. PRL 113, 082001 (2014), arXiv:1402.5175; A. Bazavov et al., Phys. Rev. D 90, 094503 (2014); Phys. Rev. D 85, 054503 (2012); Phys. Rev. D80, 014504 (2009).

[370] D. Boyanovsky, H. J. de Vega, S. Y. Wang, Nucl. Phys. A 741 (2004) 323-357, arXiv:hep-ph/0312185

[371] K. Rajagopal, F. Wilczek, Nucl.Phys. B404, 577-589 (1993), arXiv:hep-ph/9303281

[372] R. T. Andres, A. G. Nicola, arXiv:1303.6328; A. G. Nicola, R. T. Andres, Phys. Rev. D 89, 116009 (2014) arXiv:1404.2746.

[373] R. D. Pisarski, M. H. G. Tytgat, conference proceedings, arXiv:hep-ph/9705316; R. D. Pisarski, M. Tytgat, Phys. Rev. Lett. 78, 3622-3625 (1997), arXiv:hep-ph/9611206; R. D. Pisarski, M. Tytgat, summer school lecture notes, arXiv:hep-ph/9609414; R. D. Pisarski, M. Tytgat, Phys. Rev. D54, 2989-2993 (1996), arXiv:hep-ph/9604404; R. D. Pisarski, T. L. Trueman, M. H. G. Tytgat Phys. Rev. D56, 7077-7088 (1997), arXiv:hep-ph/9702362

[374] S. Jeon, J. Kapusta Phys. Rev. D54, 6475-6478 (1996), arXiv:hep-ph/9602400

[375] C. A. Dominguez, M. S. Fetea, M. Loewe, Phys. Lett. B387, 151-154 (1996), arXiv:hep-ph/9608396

[376] M. Harada, A. Shibata Phys. Rev. D55, 6716-6724 (1997), arXiv:hep-ph/9612358

[377] A. G. Nicola, D. Fernandez-Fraile, conference proceedings, arXiv:1011.3920; D. Toublan Phys.Rev. D56, 5629-5645 (1997), arXiv:hep-ph/9706273

[378] N. Fornengo, C.W. Kim and J. Song, Phys. Rev. D 56 (1997) 5123, arXiv:hep-ph/9702324; G. Mangano, G. Miele, S. Pastor and M. Peloso, Phys. Lett. B 534 (2002) 8, arXiv:astro-ph/0111408

[379] J. Bernstein, *Kinetic Theory in the Expanding Universe*, Cambridge University Press (1988)

[380] J. Lesgourgues, G. Mangano, G. Miel, S. Pastor, *Neutrino Cosmology*, Cambridge University Press (2013)

[381] F. Debbasch Physica A, 387:2443-2454 (2008).

[382] S. Tremaine and J. E. Gunn, Phys. Rev. Lett. 42, 407 (1979)

[383] S. Tremaine, M. Henon, D. Lynden-Bell, MNRAS 219, 285-97 (1986)

[384] D. Boyanovsky, H. J. de Vega, D. J. Schwarz, Ann. Rev. Nucl. Part. Sci. 56:441-500, 2006, arXiv:hep-ph/0602002

[385] J. Binney, S. Tremaine, *Galactic Dynamics*, Princeton University Press, 2008

[386] D. Boyanovsky, Phys. Rev. D77, 023528 (2008), arXiv:0711.0470

[387] J. Lesgourgues, Neutrino cosmology from (preliminary) Planck 2014, PLANCK 2014 conference in Ferrara, Italy.

[388] D. Chialva, P. S. Bhupal Dev, A. Mazumdar, Phys. Rev. D 87, 063522 (2013), arXiv:1211.0250

[389] S. Paduroiu, Y. Revaz, D. Pfenniger, arXiv:1506.03789; A. V. Maccio, S. Paduroiu, D. Anderhalden, A. Schneider, B. Moore, MNRAS,424, 1105 (2012) arXiv:1202.1282; S. Shao, L. Gao, T. Theuns, C. Frenk, MNRAS,430, 2346, (2013) arXiv:1209.5563.

[390] T. Asaka, M. Laine, M. Shaposhnikov, JHEP 0701, 091 (2007), arXiv:hep-ph/0612182; JHEP 0606, (2006) 053, arXiv:hep-ph/0605209; T. Asaka, M. Shaposhnikov, A. Kusenko, Phys.Lett.B638:401,(2006), arXiv:hep-ph/0602150; T. Asaka, M. Shaposhnikov, Phys.Lett.B620, 17 (2005), arXiv:hep-ph/0505013.

[391] J. Ghiglieri, M. Laine, JHEP 1511, 171 (2015), arXiv:1506.06752; I. Ghisoiu, M. Laine, JCAP 1412 032 (2014), arXiv:1411.1765; M. Laine, M. Shaposhnikov, JCAP 0806, 031 (2008), arXiv:0804.4543.

[392] T. Venumadhav, F.-Y. Cyr-Racine, K. N. Abazajian, C. M. Hirata, arXiv:1507.06655.

[393] C. R. Watson, Z. Li, N. K. Polley, JCAP03, 018 (2012), arXiv:1111.4217; H. Yuksel, J. F. Beacom, C. R. Watson, Phys.Rev.Lett.101, 121301 (2008), arXiv:0706.4084.

[394] M. E. Anderson, E. Churazov, J. N. Bregman, arXiv:1408.4115

[395] N. Sekiya, N. Y. Yamasaki, K. Mitsuda, arXiv:1504.02826

[396] L. Lello, D. Boyanovsky, Phys. Rev. D 91, 063502 (2015), arXiv:1411.2690

[397] Kenny C. Y. Ng, Shunsaku Horiuchi, Jennifer M. Gaskins, Miles Smith, Robert Preece, Phys. Rev. D 92, 043503 (2015), arXiv:1504.04027.

[398] A. Merle, A. Schneider, M. Totzauer, arXiv:1512.05369

[399] V. Barger, R. J. N. Phillips, S. Sarkar, Phys. Lett. B352, 365 (1995), arXiv:hep-ph/9503295.

[400] C. Dib, J. Helo, M. Hirsch, S. Kovalenko, I. Schmidt, Phys.Rev. D85 011301, (2012) arXiv:1110.5400

[401] G. Cvetic, C.S. Kim, R. Kogerler, J. Zamora-Saa, Phys. Rev. D 92, 013015 (2015), arXiv:1505.04749

[402] D. Boyanovsky, Phys. Rev. D 90, 105024 (2014) arXiv:1409.4265.

[403] Shankha Banerjee, P. S. Bhupal Dev, Alejandro Ibarra, Tanumoy Mandal, Manimala Mitra, Phys. Rev. D 92, 075002 (2015), arXiv:1503.05491

[404] J. N. Ng, A. de la Puente, B. W.-P. Pan, JHEP 1512 172 (2015), arXiv:1505.01934

[405] A. Atre, T. Han, S. Pascoli, B. Zhang, JHEP 0905, 030 (2009), arXiv:0901.3589; T. Han, B. Zhang, Phys. Lett. B97, 171804 (2006), arXiv:hep-ph/0604064; J. Gluza, T. Jelinski, Phys. Lett. B748, 125 (2015), arXiv:1504.05568; P. Coloma, B. Dobrescu, J. Lopez-Pavon, arXiv:1508.04129; P. S. Bhupal Dev, R. N. Mohapatra, arXiv:1508.02277.

[406] P. S. Bhupal Dev, A. Pilaftsis, U. Yang, Phys. Rev. Lett. 112, 081801 (2014), arXiv:1308.2209; F. F. Deppisch, P. S. Bhupal Dev, A. Pilaftsis, New J. Phys. 17, 075019 (2015), arXiv:1502.06541; A. Das, P. S. Bhupal Dev, N. Okada, Phys. Lett. B735, 364 (2014), arXiv:1405.0177; A. Das, N. Okada, Phys. Rev. D88, 113001 (2013), arXiv:1207.3734.

[407] G. Drexlin, V. Hannen, S. Mertens, C. Weinheimer, Advances in High Energy Physics, 2013 (2013), arXiv:1307.0101.

[408] S. Mertens, T. Lasserre, S. Groh, F. Glueck, A. Huber, A. W. P. Poon, M. Steidl, N. Steinbrink, C. Weinheimer, JCAP 1502, 020 (2015), arXiv:1409.0920.

[409] L. Gastaldo *et. al.* J.Low Temp.Phys. 176, 876 (2014); P. C.-O. Ranitzsch *et.al.*, arXiv:1409.0071.

[410] E. Izaguirre, G. Krnjaic, M. Pospelov, Phys. Rev. D 92, 095014 (2015), arXiv:1507.02681.

[411] K. R. Dienes, B. Thomas, Phys. Rev. D85, 083523 (2012), arXiv:1106.4546; Phys. Rev. D85, 083524 (2012), arXiv:1107.0721; K. R. Dienes, J. Kumar, B. Thomas, D. Yaylali, Phys. Rev. Lett. 114, 051301 (2015), arXiv:1406.4868.

[412] A. Boyarsky, J. Lesgourgues, O. Ruchayskiy, M. Viel, JCAP 05, 012 (2009), arXiv:0812.0010.

[413] A. Adulpravitchai, M. A. Schmidt, JHEP 1501 (2015) 006; arXiv:1507.05694.

[414] P. B. Pal, L. Wolfenstein, Phys. Rev. 25, 766 (1982).

[415] D. Notzold and G. Raffelt, Nucl. Phys. B307, 924 (1988); P. B. Pal and T. N. Pham Phys. Rev. D 40, 259 (1989).

[416] J. Wu, C.-M. Ho, D. Boyanovsky, Phys.Rev.D80:103511,(2009), arXiv:0902.4278.

[417] H. A. Weldon, Phys. Rev. D26, 1394 (1982); Phys. Rev. D26, 2789 (1982).

[418] E. Braaten, R. Pisarski, Nucl. Phys. B337, 569 (1990); Nucl. Phys. B339, 310 (1990); Phys. Rev. Lett. 64, 1338 (1990); Phys. Rev. D42, 2156 (1990); Phys. Rev. D45, 1827 (1992); Phys. Rev. D46, 1829 (1992).

[419] R. Pisarski, Nucl.Phys. A544, 527C (1992); R. Pisarski, Can.J.Phys. 71 (1993) 280 (hep-ph/9302241); R. Pisarski, Phys.Rev.Lett. 63, 1129 (1989); R. Pisarski, , Physica A158, 246 (1989)

[420] J. Frenkel, J. C. Taylor, Nucl. Phys. B334, 199 (1990).

[421] M. Le Bellac, *Thermal Field Theory*, (Cambridge Monographs on Mathematical Physics, Cambridge University Press, 1996).

[422] G. G. Raffelt, *Stars as Laboratories for Fundamental Physics*, (The University of Chicago Press, Chicago, 1996).

[423] E. Braaten, ApJ 392, 70 (1992); E. Braaten, Phys.Rev.Lett. 66 , 1655(1991); E. Braaten, D. Segel, Phys.Rev. D48, 1478 (1993) ,arXiv:hep-ph/9302213.

[424] S. Esposito, G. Mangano, G. Miele, I. Picardi, O. Pisanti, Nucl.Phys.B658, 217 (2003), arXiv:astro-ph/0301438; S. Esposito, G. Mangano, G. Miele, I. Picardi, O. Pisanti, Mod.Phys.Lett. A17 491 (2002), arXiv:astro-ph/0112384.

[425] E. W. Kolb, M. S. Turner, *The Early Universe*, (Addison Wesley, Reading, 1994).

[426] D. Boyanovsky. C.-M. Ho, Phys.Rev.D76, 085011 (2007), arXiv:0705.0703; D. Boyanovsky, New J. Phys. 17 063017, (2015), arXiv:1503.00156.

[427] D. Boyanovsky, J. Wu, Phys.Rev.D83, 043524 (2011), arXiv:1008.0992.

[428] A. Palazzo, D. Cumberbatch, A. Slosar, J. Silk, Phys. Rev.D 76, 103511 (2007), arXiv:0707.1495

[429] B. Brandt, A. Francis, H. Meyer, D. Robaina, arXiv:1506.05732

[430] J. Rafelski, arXiv:1211.4297

[431] D. Boyanovsky, L. Lello, New Journal of Physics 16, 063050 (2014), arXiv:1403.6366; D. Boyanovsky, Nucl. Phys. B888, 248 (2014), arXiv:1406.5739.

[432] R. Cicchetti, A. Faraone, IEEE Trans. Antennas and Propagation 52 (12) p.3373-3389 (2004).

[433] A. D. Dolgov, S. H. Hansen, Astroparticle Physics 16, 339 (2002), arXiv:hep-ph/0009083.

[434] M. Drewes, J. U. Kang, JHEP 1605 (2016) 051, arXiv:1510.05646.

[435] W. Bonivento *et.al.* arXiv:1310.1762.

[436] G. Gelmini, E. Osoba,S. Palomares-Ruiz, S. Pascoli, JCAP 10,029 (2008),arXiv:0803.2735.

[437] H. Ishida, M. Kusakabe, H. Okada, Phys. Rev. D90, 083519 (2014), arXiv:1403.5995.

[438] For a review see: B. D. Fields, Annual Review of Nuclear and Particle Science, 61, 47-68 (2011); arXiv:1203.3551.

[439] F. Iocco, G. Mangano, G. Miele, O. Pisanti, P. D. Serpico, Phys. Rept. 472, 1 (2009), arXiv:0809.0631.

[440] M. Pospelov,J. Pradler, Ann. Rev. Nucl. Part. Sci. 60, 539 (2010), arXiv:1011.1054.

[441] V. Poulin, P. D. Serpico, Phys. Rev. Lett. 114, 091101 (2015), arXiv:1502.01250; Phys. Rev. D 91, 103007 (2015), arXiv:1503.04852.

[442] L. Salvati, L. Pagano, M. Lattanzi, M.Gerbino, A. Melchiorri, arXiv:1606.06968.

[443] S.H. Hansen and Z. Haiman, Astroph. Jour. 600, 26 (2004), arXiv:astro-ph/0305126.

[444] O. Ruchayskiy, A. Ivashko, JHEP 1206 (2012) 100, arXiv:1112.3319.

[445] A. C. Vincent, E. Fernandez- Martinez, P. Hernandez, M. Lattanzi, O. Mena, JCAP 04 (2015) 006, arXiv:1408.1956.

[446] M. Drewes, arXiv:1510.07883, M. Drewes, B, Garbrecht, arXiv:1502.00477.

[447] L. Lello, D. Boyanovsky, JCAP06(2016)011, arXiv:1508.04077.

[448] M. Laine, M. Meyer, JCAP 1507 035 (2015), arXiv:1503.04935.

[449] C. M. Ho, D. Boyanovsky, H. J. de Vega, Phys.Rev. D72 085016 (2005),arXiv:hep-ph/0508294.

[450] D. Boyanovsky, K. Davey, C. M. Ho, Phys.Rev. D71 023523 (2005), arXiv:hep-ph/0411042.

[451] D. Boyanovsky, New J. Phys. 17 063017 (2015), arXiv:1503.00156.

[452] D. Gorbunov, M. Shaposhnikov, JHEP0710:015 (2007), arXiv:0705.1729

[453] B. Audren, J. Lesgourgues, G. Mangano, P. D. Serpico, T. Tram, JCAP 12 (2014) 028, arXiv:1407.2418.

[454] V. Poulin, P. D. Serpico, J. Lesgourgues, arXiv:1606.02073.

[455] R. R. Volkas and Y. Y. Y. Wong, Phys. Rev. D 62, 093024 (2000), arXiv:hep-ph/0007185; K. S. M. Lee, R. R. Volkas, and Y. Y. Y. Wong, Phys. Rev. D 62, 093025 (2000), arXiv:hep-ph/0007186.